Aggregate Outcomes of Nonlinear Prices in Supply Chains*

Luca Lorenzini
UCLA Anderson

Antonio Martner
UCLA and Central Bank of Chile

Job Market Paper

This version: October 24, 2025
Link to last version

Abstract

We study the welfare implications of nonlinear pricing in supply chains. Using population-level firm-to-firm transactions from Chile, we document quantity-dependent and buyer-specific pricing strategies. We develop a general equilibrium model where firms pay and charge nonlinear prices as two-part tariffs—a flat fee and a marginal price—that match observed price patterns. Relative to a uniform pricing interpretation of the same data, nonlinear pricing increases output per firm but distorts firm entry because flat fees redistribute profits unevenly across firms. We quantify the model and find that welfare under nonlinear prices attains about 75% of the efficient benchmark. When we study a counterfactual policy banning all price discrimination—constraining firms to uniform pricing—welfare falls to about 49% of the efficient benchmark. Firms constrained to uniform pricing raise marginal prices that compound along supply chains, amplifying deadweight losses through markup accumulation. When interpreting the same data as uniform rather than nonlinear pricing, measured welfare is 57% versus 75% of the efficient benchmark. These results indicate that the measured aggregate welfare impact of market power depends meaningfully on the extent to which firms use nonlinear pricing.

*We are deeply indebted to Hugo Hopenhayn for his insights and constant encouragement. Antonio would like to thank Michael Rubens, John Asker, and Ariel Burstein for invaluable mentorship and support. Jonathan Vogel, David Baqaee, Lee Ohanian, Federico Huneeus, Yasutaka Koike-Mori and Mounu Prem provided valuable comments. The views expressed are those of the authors and do not necessarily represent the views of the Central Bank of Chile or its board members. Authors' email: lucalorenzini@ucla.edu, amartner@ucla.edu

1 Introduction

Price discrimination in supply chains is central to current debates on market power and antitrust policy. While price discrimination has long been recognized as pervasive in firms' pricing strategies—"one of the most prevalent forms of marketing practices" (Varian, 1989)—renewed antitrust scrutiny now targets these practices in firm-to-firm transactions.[1] Despite recognition and policy attention, population-scale evidence on price discrimination along supply chains and its implications for resource allocations, rent sharing, and aggregate welfare remains scarce.

Using population-level administrative data on firm-to-firm transactions in Chile, we find indicative evidence of widespread price discrimination in supply chains: unit prices decrease with quantity purchased and vary systematically across buyer sectors, consistent with buyer-sector-specific nonlinear price schedules. These patterns depart from the uniform-pricing assumption standard in the literature—a single, quantity-invariant price for all buyers—and point instead to nonlinear pricing with buyer-sector-specific schedules as the prevalent pricing strategy in supply chains.

Guided by these patterns, we develop a multi-sector general equilibrium model in which firms simultaneously charge and pay nonlinear prices. We show that, under standard assumptions, the optimal contract takes the form of a buyer-industry-specific two-part tariff—a flat fee and a marginal price. Relative to uniform pricing, nonlinear pricing brings quantities closer to efficient levels through lower marginal prices, mitigating markup accumulation along the supply chain and attenuating double marginalization. While improving allocative efficiency, the flat fees introduce new distortions through rent extraction, affecting profit distributions and entry decisions in theoretically ambiguous ways.

We calibrate the model's parameters and validate that it closely replicates the nonlinear price schedules across buyer sectors observed in the data. We then quantify two questions. First, what are the aggregate welfare implications of nonlinear pricing? Interpreting observed prices as nonlinear, we find that welfare attains about 75% of the efficient benchmark. When we study a counterfactual policy banning all price discrimination—constraining firms to uniform pricing—welfare falls to about 49% of the efficient benchmark. These losses arise primarily from worsened allocative efficiency: constrained to uniform pricing, firms cannot extract rents without distorting quantities, so they raise marginal prices that compound along the supply chain, amplifying double marginalization.

Second, how does the measurement of the aggregate welfare costs of market power depend on the assumptions about firms' pricing behavior? When the same data is interpreted as uniform prices, measured welfare attains 57% of the efficient benchmark. The welfare gap mirrors that from banning price discrimination—a methodological point with substantive implications: ignor-

[1] In *FTC v.* Southern Glazer's (dec 2024), the complaint alleges discriminatory quantity discounts and rebates that are inaccessible to smaller rivals and not justified by cost.

ing firms' ability to price discriminate can substantially overstate the aggregate costs of market power.

We start by studying the canonical screening framework in partial equilibrium, where a seller faces buyers with privately observed scale types and offers a menu of contracts.[2] We consider a monopolist with constant marginal cost selling to heterogeneous buyers[3]. When buyer types are Pareto distributed and buyers' revenue functions are homogeneous under isoelastic inverse demand, the optimal nonlinear pricing schedule takes the form of a two-part tariff: a constant marginal price and a flat fee. The marginal price governs allocation while the flat fee extracts rents (e.g., Wilson, 1993; Armstrong, 1996), reshaping profits across firms without affecting allocations. While entry is held fixed in this partial-equilibrium framework, the same mechanism, if extended to general equilibrium, would influence firms' profitability and thereby their entry incentives.

We show that this structure brings marginal prices closer to marginal costs relative to uniform pricing, improving allocative efficiency, while flat fees redistribute surplus without affecting input choices. The two-part tariff yields a sharp empirical prediction: average unit prices decline with purchase size and converge asymptotically to a common marginal price. Hence, observed unit prices may not be entirely allocative, containing a redistributive component.

To test these theoretical predictions, we analyze Chile's population of firm-to-firm invoices. Within each seller-product pair, unit prices fall with quantity and flatten at higher quantities—a pattern pervasive across seller industries that shifts systematically with buyer industries. This evidence reveals a combined pricing scheme: buyer-sector specific nonlinear prices that exhibit curvature within sectors (second-degree price discrimination) while shifting in level across sectors (third-degree price discrimination). The patterns align with our two-part tariff prediction, where average unit prices decline with quantity and converge toward a common marginal price. While we present evidence inconsistent with several alternative explanations, we maintain a conservative interpretation: the observed price–quantity relationships are indicative of price discrimination. Rather than using them to discipline the model, we leave them as untargeted moments and assess how much a model calibrated to the firm size distribution can account for the observed quantity discounts.

To quantify the welfare implications of these price discrimination patterns, we build a multi-sector general equilibrium supply chain model where firms endogenously and simultaneously charge and pay nonlinear prices. As in the partial equilibrium framework, we assume Pareto-distributed firm productivity; through our technology specifications, we endogenously obtain isoelastic demands and constant marginal costs. Although each assumption has antecedents in prior work, their joint adoption is central to our contribution: together they replicate the empirical patterns of quantity discounts while preserving tractability in a general-equilibrium supply-

[2]See Mirrlees, 1971; Mussa and Rosen, 1978; Maskin and Riley, 1984; Tirole, 1988, ch. 3.

[3]In canonical screening models, buyer surplus corresponds to utility. Because buyers in our setting are firms, their surplus is instead represented by total revenues, which play an analogous role.

chain setting where firms simultaneously charge and pay nonlinear prices. Sellers implement both second-degree price discrimination (quantity-based menus within sectors) and third-degree price discrimination (buyer-sector-specific schedules). We highlight two theoretical results. First, the optimal contract remains a sector-specific two-part tariff with constant marginal price and flat fee. Second, the framework accommodates arbitrary firm linkages with cycles, where each firm simultaneously charges and pays nonlinear prices.

We benchmark the nonlinear pricing equilibrium against the planner's first-best allocation—achieved in a decentralized equilibrium through a ban on price discrimination combined with an output subsidy that restores marginal-cost pricing conditional on entry (Baqaee and Farhi, 2020a). The model enables an exact welfare decomposition using sufficient statistics: sector-specific final demand exposures, markups, and firm masses. We decompose welfare into an intensive margin (allocative efficiency from quantity distortions) and an extensive margin (variety from entry distortions). This decomposition reveals how sector-level markups translate into aggregate welfare losses through both direct and indirect final-demand exposure via input-output linkages. Crucially, under nonlinear pricing, the relevant markup for allocative efficiency is the marginal price markup rather than the average markup, while flat fees affect welfare only through entry.

We calibrate the model using the population of firm-to-firm transactions in Chile, merged with firms' accounting balance sheets, and estimate substitution elasticities from a quasi-experimental price shock. We solve for equilibrium under three scenarios: (i) nonlinear pricing, (ii) uniform pricing, and (iii) the planner's first best. Two parameterization strategies guide the analysis. First, we estimate parameters under the nonlinear pricing interpretation and hold them fixed in our policy counterfactual, assessing how welfare changes when price discrimination is banned. Second, we fully recalibrate the model using the same data but interpreting observed prices as uniform—quantity-invariant charges identical across buyers. This dual approach allows us to address both our policy question (welfare effects of banning discrimination) and our measurement question (how pricing interpretations affect estimated welfare costs of market power).

The model calibrated under nonlinear pricing replicates both the magnitude and curvature of observed price-quantity relationships. Under price discrimination, welfare attains 75% of the efficient benchmark, with the intensive margin accounting for 79% of the welfare gap. A policy banning all price discrimination reduces welfare from 75% to 49% of the efficient benchmark—equivalently, the efficiency shortfall widens from 25% to 51%, roughly doubling. The welfare losses concentrate in upstream sectors with high final demand exposure, particularly Construction and Retail & Wholesale, and are driven primarily by the intensive margin. When we recalibrate the model interpreting the same data as uniform pricing, welfare attains 57% of the efficient benchmark, again driven mainly by the intensive margin.

Our results highlight three key messages. First, price discrimination is pervasive in firm-to-firm transactions and has first-order implications for assessing the aggregate outcomes of market

power. Second, prohibiting all forms of price discrimination can generate substantial welfare losses. When firms are constrained to uniform prices and cannot extract rents without distorting quantities, they raise marginal prices that compound allocative inefficiencies along the supply chain. A ban on price discrimination is therefore incomplete if not complemented by output subsidies and may even be counterproductive. Third, aggregate welfare costs of market power may be substantially overstated when researchers assume uniform pricing despite evidence of pervasive nonlinear pricing.

Related Literature. This paper relates to three strands of literature. First, price discrimination and screening. We provide population-scale evidence on quantity-dependent pricing and buyer-specific schedules in firm-to-firm transactions, embedding these practices in a general equilibrium framework to quantify their aggregate implications. Our empirical findings relate to documented input price dispersion across buyers and its allocative consequences (Dhyne et al., 2023; Burstein et al., 2024). Theoretically, our model builds on classic price discrimination and screening results (Dupuit, 1844; Mussa and Rosen, 1978; Maskin and Riley, 1984; Wilson, 1993) and price discrimination surveys under imperfect competition (Varian, 1989; Stole, 2007). We contribute by quantifying how nonlinear prices reshape effective marginal prices, pass-through, and markup accumulation along supply chains in general equilibrium, where firms simultaneously pay and charge nonlinear prices.

Second, production networks and misallocation. We show how distortions propagate through input-output structures and affect aggregate outcomes, building from Quesnay (1894)'s "Tableau économique" to modern developments (Restuccia and Rogerson, 2008; Hsieh and Klenow, 2009; Jones, 2011; Oberfield, 2018; Carvalho and Tahbaz-Salehi, 2019; Bigio and La'o, 2020; Baqaee and Farhi, 2020b). While contracting frictions distort input choices and firm organization (Boehm and Oberfield, 2020), and network topology amplifies wedges into macroeconomic outcomes (Jones, 2011), we reinterpret markup dispersion and buyer-specific pricing as endogenous outcomes of price discrimination. Nonlinear pricing changes which margins are distorted—entry versus production, average versus marginal prices—thus affecting how misallocation is measured in general equilibrium. Our approach builds on aggregation and welfare accounting with heterogeneity and linkages (Baqaee and Farhi, 2020b) and complements micro evidence on buyer-level price dispersion (Burstein et al., 2024).

Third, the aggregate welfare costs of market power, starting from Harberger (1954) and extended by (Hall, 2018; Barkai, 2020; Autor et al., 2020; De Loecker et al., 2020). In dynamic heterogeneous-firm models, markups entail large welfare costs and interact with entry (Edmond et al., 2023). We show that observed price discrimination—pervasive in supply-chain data (Dhyne et al., 2022; Burstein et al., 2024)—can mitigate allocative inefficiency relative to uniform pricing, even as it raises surplus extraction.

This echoes general-equilibrium results showing that nonlinear pricing changes the interpretation of markup heterogeneity and introduces buyer-side misallocation, as in Bornstein and Peter (2024). They study retail markets where firms offer a single nonlinear schedule to consumers, showing that nonlinear pricing shifts quantities toward high-taste buyers and away from low-taste ones, thereby amplifying consumer-firm wedges and reducing welfare relative to linear pricing.

Our analysis differs both in question and in framework. We study input-market contracts within production networks with free entry, where firms simultaneously charge and pay nonlinear prices. Modeling both sides of the supply chain requires a new structure—one in which sector-specific nonlinear prices combine elements of second- and third-degree price discrimination. Our quantitative exercise assume uniform and nonlinear pricing as alternative interpretations of the same data: holding technology and demand primitives fixed, we re-estimate pricing parameters conditional on that interpretation. This design enables commensurate welfare comparisons and shows that measured losses from market power depend critically on the assumed pricing conduct.

2 Optimal Nonlinear Price Characterization in Partial Equilibrium

We present a framework of optimal nonlinear pricing in partial equilibrium, deriving a result that serves as the building block for a general equilibrium supply chain model, in which firms charge and pay nonlinear prices. We find that under Pareto-distributed buyer types, homogenous revenue functions, and constant marginal costs, the optimal price schedule takes the form of a two-part tariff: a constant per-unit price and a flat fee as shown by Laffont and Tirole (1993), building on Spence (1977) and Maskin and Riley (1984).

We use the canonical monopolistic screening problem (Tirole, 1988). A seller with constant marginal cost $c > 0$ faces a continuum of buyers with privately observed productivity types $z \in [\underline{z}, \infty)$, drawn from $F(z)$ with density $f(z)$. The seller chooses a menu, that is, a pair of measurable functions $(q, T) : [\underline{z}, \infty) \to \mathbb{R}_+ \times \mathbb{R}$, which jointly assign to each type z a quantity $q(z)$ and a transfer (i.e., total payment) $T(z)$. Under nonlinear pricing, the transfer $T(z)$ need not equal price times quantity, as under uniform linear pricing, where it would be $p\,q(z)$, and may include, for example, a fixed fee and a per-unit component. Given q, buyer z generates revenue[4] $R(z, q)$ and obtains net surplus $\Pi(z) = R(z, q(z)) - T(z)$. The seller's problem is

$$\max_{(q,T)} \ \Pi_{\text{seller}} = \int_{\underline{z}}^{\infty} \Big[T(z) - c\,q(z)\Big] f(z)\,dz \tag{1}$$

$$\text{s.t.} \quad (IR) \quad \Pi(z) \equiv R(z, q(z)) - T(z) \geq 0,$$

$$(IC) \quad \Pi(z) \geq R\Big(z, q(\tilde{z})\Big) - T(\tilde{z}), \qquad \forall\, z, \tilde{z} \in [\underline{z}, \infty)$$

[4]This corresponds to the buyer's utility in the canonical screening framework. Because our buyers are firms, we interpret it as revenue rather than utility.

Individual rationality (IR) and incentive compatibility (IC) ensure participation and truthful revelation. We normalize the outside option to zero, so IR is $\Pi(z) \geq 0$. In the optimum, the lowest-type IR binds ($\Pi(\underline{z}) = 0$). Under standard single-crossing and concavity conditions, as shown in Appendix A, the seller's problem can be rewritten as a pointwise optimization[5]:

$$\max_{\{q(z)\}} \Pi_{\text{seller}} = \int_{\underline{z}}^{\infty} \Big[\phi(z, q(z)) - c\, q(z)\Big] f(z)\, dz$$
$$\text{with} \quad \phi(z,q) \;=\; R(z,q) \;-\; \frac{1}{h(z)} \frac{\partial R(z,q)}{\partial z}, \qquad h(z) \;=\; \frac{f(z)}{1 - F(z)}$$

where $\phi(z, q)$ is the virtual surplus and $h(z)$ is the hazard rate. The virtual surplus, which represents the seller's effective revenue from serving type z, consists of two components. The first term, $R(z, q)$, is buyer z total revenue. If the monopolist could price-discriminate perfectly, this would coincide with the seller's revenue as well. The second term, $-\frac{1}{h(z)} \frac{\partial R(z,q)}{\partial z}$, captures the truth-telling cost, i.e., the additional rents the seller must leave to higher types to prevent them from mimicking type z.

The hazard rate summarizes how many buyers lie above z [i.e., $1 - F(z)$] relative to the density at z [i.e., $f(z)$], and thus measures how easy it is to enforce truth-telling. Taking the FOC for buyer type z,

$$\frac{\partial}{\partial q}\Big[\phi(z, q(z)) - c\, q(z)\Big] = 0 \quad \Longrightarrow \quad \phi_q(z, q(z)) = c, \quad \text{hence:}$$
$$R_q(z, q(z)) \;=\; c \;+\; \frac{1}{h(z)} R_{zq}(z, q(z)) \tag{2}$$

Increasing q for type z yields direct marginal revenue $R_q(z, q(z))$, but it also raises the truth-telling cost by $\frac{1}{h(z)} R_{zq}(z, q(z))$ (the extra rents that must be left to higher types). The optimal contract sets marginal virtual revenue equal to marginal cost.

In addition to constant marginal cost, we impose two further assumptions. First, buyer types are distributed according to a Pareto distribution with tail parameter κ. Second, buyers' revenue functions are homogeneous on the quantity transacted with the seller, so buyer type shifts demand for the seller's good without altering its curvature. Specifically,

$$R(z, q) = z^{\frac{\sigma-1}{\sigma}}\, q^{\frac{\sigma-1}{\sigma}}, \qquad \sigma > 1 \tag{3}$$

where σ is the curvature parameter (the demand elasticity faced by the seller). These two assumptions imply the tail condition $\kappa > \sigma - 1$, which guarantees that total sales are finite.

[5] Assuming monotonicity in the allocation. Because of the latter functional form's assumptions, ironing is unnecessary.

Lemma 1 (Optimal two-part tariff under homogeneous revenue function and Pareto types[6]). *Consider the screening in Equation 1. Suppose (i) revenue is homogeneous in quantity as in Equation 3, with shape parameter $\sigma > 1$, and (ii) buyer types are Pareto-distributed with tail parameter $\kappa > \sigma - 1$ with lower type $\underline{z}$, so that $h(z) = \kappa/z$. Under these assumptions, the optimal nonlinear price schedule takes the form of a two-part tariff $\{F, p^{\text{NLP}}\}$:*

$$T(z) = F + p^{\text{NLP}}\, q(z), \qquad \textit{where:} \quad p^{\text{NLP}} = \frac{\rho}{\rho - 1}\, c, \qquad \rho \equiv \frac{\kappa\, \sigma}{\sigma - 1} > 1, \qquad F:\ \Pi(\underline{z}) = 0$$

F is a flat fee chosen so that the lowest-served type's participation constraint binds, $\Pi(\underline{z}) = 0$. We refer to p^{NLP} as the marginal price, to distinguish it from the unit price, since it applies only to the incremental quantity purchased.[7]

Three remarks follow. First, under the Pareto distribution of types, the virtual surplus at the lower bound is strictly positive. Hence, the integrand $[\phi(z, q(z)) - c\, q(z)]\, f(z)$ is positive in a neighborhood of $\underline{z}$, so excluding any mass of low types strictly reduces profit by the foregone positive contribution (see Appendix A.1). Hypothetically, excluding the lowest type $\underline{z}$ would allow the monopolist to raise the flat fee, but at the cost of losing demand from $\underline{z}$. Because a Pareto distribution places a large mass of buyers near the bottom, each excluded buyer contributes little individually, but many are lost at once, making the demand loss larger than any additional flat-fee revenue from those who remain. Therefore, exclusion is never optimal.

Second, the quantities allocated $q(z_i)$ are determined by the marginal price p^{NLP}; the flat fee F only redistributes surplus but does not change q. Seller profit from transacting with type z_i has two components: a variable-profit rectangle $(p^{\text{NLP}} - c)\, q(z_i)$ and a flat-fee component F, which is set by the lowest-type participation constraint (Figure 1a). The deadweight loss is the area between the demand curve and marginal cost c, over the range of quantities from q^{NLP} to the efficient level q^*. Changing F does not affect this area, while changing the marginal price does.

Third, the flat fee F is identical across buyer types and purely redistributes surplus. Spreading this fixed amount over more units makes the average unit price fall with quantity: $T(z)/q(z) = p^{\text{NLP}} + F/q(z)$ (Figure 1b). For small purchases, the flat-fee share $F/q(z)$ is large and the average unit price sits well above the allocative marginal price; as quantity grows, $F/q(z)$ becomes negligible and the average unit price converges to the allocative marginal price p^{NLP}, which governs quantities.

We define the total markup as the ratio of the average unit price to marginal cost. We then decompose this markup into two components. The first component, which we call the allocative

[6] We get to the same solution using the Wilson (1993) approach on demand profiles, as shown in Appendix A.

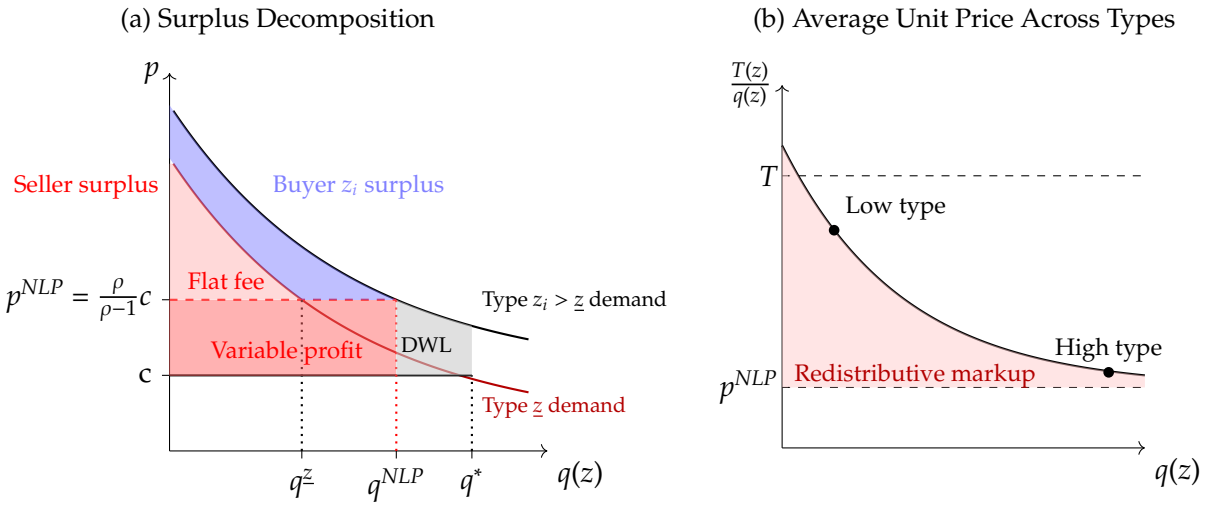
[7] Proof sketch. Substituting $h(z) = \kappa/z$ and homogeneous $R(z, q)$ into the first-order condition (Equation (2)) yields a constant marginal price that solves $R_q(z, q) = \frac{\rho}{\rho - 1} c$ with $\rho = \kappa\sigma/(\sigma - 1)$. The envelope condition and IC then pin down transfers up to a constant; choosing F to satisfy $\Pi(\underline{z}) = 0$ completes the two-part tariff. Full derivations are provided in Appendix A.

markup, is given by the ratio of the marginal price to marginal cost and equals $\frac{\rho}{\rho-1}$. We refer to it as allocative because it alters the quantity allocated, generating in this case a deadweight loss (Figure 1a). The second component is the redistributive markup, which does not affect the allocation but instead redistributes surplus from buyer to seller:

$$\underbrace{\mu_{\text{tot}}}_{\text{Total Markup}} := \frac{T(z)/q(z)}{c} = \underbrace{\frac{\rho}{\rho-1}}_{\text{Allocative Markup}} + \underbrace{\frac{F}{q(z)c}}_{\text{Redistributive Markup}}$$

It follows that the total markup paid decreases with the quantity purchased, $q(z)$, as illustrated in Figure 1b.

Figure 1: Surplus and Unit Average Prices



Notes. Panel A: the per-unit price pins down quantity; the flat fee is a lump sum that redistributes surplus without affecting q. Markup revenue is the rectangle $(p^{\text{NLP}} - c)q^{\text{NLP}}$; efficient and two-part-tariff quantities are labeled q^* and q^{NLP}. Panel B: the average unit price $T(z)/q(z) = F/q(z) + p^{\text{NLP}}$ is higher for low types and declines with z toward p^{NLP}. The per-unit flat fee is larger for low types and fades with type.

Implementability in supply chains and testable prediction. In Section 4, we build a multisector general equilibrium framework. Each seller trades with buyers from multiple sectors, and buyers within each sector are heterogeneous in productivity. Firms both pay nonlinear prices upstream and charge nonlinear prices downstream. As a result, revenue and marginal-cost functions—which in the previous section we treated as primitive—become endogenous, general-equilibrium objects shaped by nonlinear pricing. We assume that sellers can discriminate across sectors but not across types within a sector. Our main result is that, when firm heterogeneity within a sector follows a Pareto distribution, the equilibrium nonlinear contracts take the form of two-part tariffs, as in Lemma 1. Specifically, we show that the allocative markup and flat fee are seller-sector

specific: they are identical across buyers within a sector but vary across sectors.

This characterization delivers a testable prediction: if pricing is equivalent to a two-part tariff, total payment $T = F + pq$ implies an average unit price $T/q = F/q + p$ that is strictly decreasing and convex in q, with a horizontal asymptote at p (Figure 1b). In the next section, using product-level buyer-seller transaction records from Chile, we test for departures from uniform pricing by examining whether average unit prices display this pattern.

3 Evidence on Nonlinear Pricing in Supply Chains

We document the presence and shape of nonlinear prices using the population of firm-to-firm transactions in Chile. Three main findings emerge. First, unit prices vary with quantity transacted and buyer characteristics, inconsistent with uniform pricing. Second, this variation is well-approximated by two-part tariffs: average unit prices fall with quantity while marginal prices converge to a constant. Third, the steepness of these schedules differs across seller and buyer industries, indicating heterogeneity in the strength of nonlinear prices.

Data description. We use data from the population of Chilean firm-to-firm value-added tax invoices collected by the Chilean Internal Revenue Service.[8] For each transaction-specific invoice, we observe seller and buyer IDs, a free-text product "detail," and the corresponding price and quantity. These transaction records can be merged with firms' accounting variables, including total revenue, employee headcounts, labor costs, materials costs, and capital expenditure.

We work at the most granular level and keep the full economy-wide universe of transactions available for 2024, without industry exclusions. Our unit of observation is each invoice line item between two tax identifiers.[9] The "detail" field is often seller-specific (e.g., blue paint, brand XX, 3 gallons), so we treat products as seller-product pairs. In most transactions, shipping appears as a separate line, so unit prices exclude shipping.[10] Our approach complements Burstein et al. (2024), who use the same administrative source and document important price-dispersion facts in a manufacturing subsample; here, we exploit the complete data available across all industries and

[8]This study was developed within the scope of the research agenda conducted by the Central Bank of Chile (CBC) in economic and financial affairs of its competence. The CBC has access to anonymized information from various public and private entities, by virtue of collaboration agreements signed with these institutions. To secure the privacy of workers and firms, the CBC mandates that the development, extraction, and publication of the results should not allow the identification, directly or indirectly, of natural or legal persons. Officials of the CBC processed the disaggregated data. All the analysis was implemented by the authors and neither involved nor compromised the Chilean IRS. The information contained in the databases of the Chilean IRS is of a tax nature originating in self-declarations of taxpayers presented to the Service; therefore, the veracity of the data is not the responsibility of the Service.

[9]Not necessarily firms, as some tax IDs do not report hiring workers, purchasing intermediate inputs, or capital expenditure.

[10]Including shipping could generate declining average unit prices with quantity, which would reflect scale economies in shipping rather than nonlinear pricing contracts. By excluding shipping charges, we ensure that variation in average unit prices reflects contractual form rather than transportation technology.

retain maximum granularity to study nonlinear pricing in supply chains.

We perform three minimal data-cleaning steps to limit measurement error. First, we keep transactions with positive prices and quantities and nonmissing product detail. Second, we keep firms that reported positive sales in at least one month during 2024. Third, to avoid spurious variance, we drop products with at least two transactions where the same-day max-to-min price ratio exceeds the 99th percentile of its daily distribution. These filters retain 98% of the transactions. The final sample contains 537,521 seller IDs and 3,398,323 buyer IDs that traded 60,029,741 distinct products across 1.24 billion transactions in 2024.

Price determinants. We begin by quantifying within-seller-product price dispersion. None of the exercises in this section aim to be causal, but rather they describe equilibrium objects observed in the data and test which variables they correlate with in search of indicative evidence. We observe substantial price variations for a given seller i and product g (the "detail" variable in the invoice) within a month. Following Burstein et al. (2024), we construct a price-dispersion measure $\tilde{p}_{igt}$ for June 2024 (t = month), the month with the most transactions in 2024. We divide unit prices observed for each product g transaction from seller i to buyer j by the mean price across seller i and product g and month m. We repeat the same exercise for June 19, 2024, (t =day) the day with the most transactions that month, to ensure that our results are not driven by month-specific demand and supply shocks. The variance of $\ln \tilde{p}_{igt}$ is 0.65 monthly and 0.61 daily, and around 30% of transactions in both cases show no price dispersion.

Figure 2 displays $\tilde{p}_{igt}$ histograms. The histograms are similar when computed at daily versus monthly frequency, suggesting that residual price variation is not primarily driven by supply or demand shocks, nor by inflation. For 71% of transactions, we cannot reject that firms engage in some form of price discrimination departing from uniform pricing.

These facts motivate a decomposition of the residual dispersion into quantity versus buyer components. This approach will be nonparametrical where each different quantity transacted and buyer firm ID is assigned a specific dummy variable that will be contained in the fixed effects. To net out common shocks, we first estimate

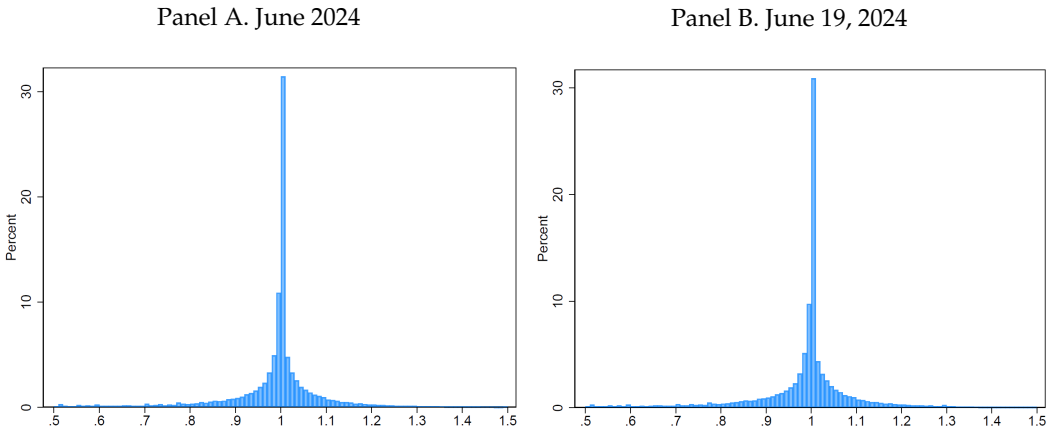
$$\ln p_{igjt} = \beta_0 + \Psi_{igd} + \epsilon_{ijgt}$$

where Ψ_{igd} are seller-product-day fixed effects and t indexes the time stamp during the day. The residual ϵ_{ijgt} captures price differences across buyers of the same (i, g) on the same day. We then project ϵ_{ijgt} using alternative fixed-effect sets S:

$$\epsilon_{ijgt} = \beta_0 + \Psi_S + \nu_{ijgt} \tag{4}$$

where Ψ_S includes (i) functions of transaction quantity, (ii) buyer-group fixed effects (sector-size-

Figure 2: Price Dispersion



Notes: This figure reports the distribution of the log of demeaned price for the month of June 2024. We exclude seller-product pairs with only one transaction.

region; 626 groups), and (iii) interactions of quantity with buyer-group to allow group-specific discount schedules. This two-step procedure yields an R^2-based horse race over residual price variation. The results provide indicative evidence on the importance of second-degree (quantity) and third-degree (buyer) components. Identification nuances in separating these mechanisms, and additional robustness using monthly fixed effects and industry subsamples, are detailed in Appendix B.1.

Table 1 reports the R^2 across specifications. Quantity alone explains 34% of residual price variation (column 1), indicating that quantity discounts play an important role in explaining residual price variation. Coarser buyer-group effects (sector-size-region) still account for 28% (column 2), indicating that most of the variance explained by buyer fixed effects is captured by observable group characteristics. Allowing group-specific discount schedules (quantity-buyer-group) explains 53% (column 3), consistent with hybrid second- and third-degree price discrimination accounting for the lion's share of price dispersion along supply chains.[11]

[11]We omitted buyer fixed effects alone because of absence of price variation by buyer and day for the same seller and product. Appendix B.1 includes them at the monthly level, indicating stable heterogeneity across buyers, but never generating a higher R^2 relative to quantities and buyer groups interacted.

Table 1: Price Residual Determinants

	(1)	(2)	(4)
R^2	0.344	0.275	0.535
S = Quantity	✓		
S = Buyer Group		✓	
S = Quantity × Buyer group			✓
N	147M	147M	147M

Notes: This table reports R^2 values from regressions of price residuals ϵ_{ijgt} on different specifications S, where residuals are obtained from Equation 4 after controlling for seller-product-day fixed effects. Buyer groups are defined by combinations of 11 sectors, 3 size categories, and 16 regions.

Nonlinear prices. We test for nonlinear pricing by examining whether observed unit prices (which we interpret in equilibrium) covary systematically with transaction quantities. We estimate

$$\ln p_{igjt} \;=\; \beta_1 \ln q_{igjt} \;+\; \Psi_{igd} \;+\; \Psi_S \;+\; \varepsilon_{igjt} \tag{5}$$

where p_{igjt} and q_{igjt} are the unit price and quantity for seller i, product g, buyer j, at transaction day t on day d. Ψ_{igd} are seller-product-day fixed effects; Ψ_S varies by specification to add buyer or buyer-group controls and their interactions (Table 2). A potential concern for interpreting β_1 as evidence of price discrimination is that supply shocks could simultaneously reduce prices and raise quantities, creating spurious correlation. Because we condition on seller-product-day fixed effects, identification comes only from within-seller-product-day price variation, making this interpretation unlikely. Another concern is that some buyers may systematically purchase larger quantities and also obtain lower prices due to monopsony power; by including buyer or buyer-group controls in Ψ_S, we assess how much of the observed variation can be explained by this mechanism. We estimate (5) on the universe of 2024 transactions after dropping singletons.

Column 1 conditions on seller-product-day and yields a quantity coefficient of -0.042 log points, so doubling quantity is associated with a 2.9% ($\beta_1 \ln 2$) lower unit price. Adding buyer fixed effects in column 2 strengthens the coefficient to -0.084, indicating that once persistent buyer heterogeneity is absorbed, quantity discounts are even more pronounced. Replacing buyer FE with buyer-group FE (sector-size-region) still gives a sizable -0.065 in column 3. Column 4 allows fully flexible group-specific schedules by interacting Ψ_{igd} with buyer group; the coefficient remains negative and significant at -0.037, about 90% of the column 1 magnitude, consistent with systematic quantity discounts across buyer groups.[12]

[12]We do not interact quantity with buyer fixed effects. Within a day for a given seller-product, the same buyer rarely purchases multiple distinct quantities; moreover, such a specification would push toward buyer-specific nonlinearities closer to first-degree discrimination, which we view as implausible in this setting.

Table 2: Price-Quantity Coefficient Estimates

	(1)	(2)	(3)	(4)
$\ln q_{igjt}$	-0.042	-0.084	-0.065	-0.037
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
S_{Base} = Seller × Product × Day	✓			
$S = S_{Base}$+ Buyer		✓		
$S = S_{Base}$+ Buyer Group			✓	
$S = S_{Base}$ × Buyer Group				✓
N	430M	430M	430M	430M
R^2	0.9646	0.9678	0.9659	0.9790

Notes: This table reports coefficients from regressions of log unit prices on log quantities with varying fixed-effect specifications S. *Base* refers to seller-product-day fixed effects. Buyer groups are defined by combinations of 11 sectors, 3 size categories, and 16 regions. Standard errors are in parentheses. All regressions use the universe of Chilean firm-to-firm transactions in 2024 after dropping singletons.

We repeat the same exercise from column 1 for each 1-digit sector in the economy; we show the results in Appendix B.2. We find that the smallest-quantity coefficient is around 0.00 in utilities, while the largest is observed in construction, at 0.13.

Could buyer monopsony power drive the price-quantity correlation? Table 2 argues against it: adding buyer fixed effects strengthens the coefficient from -0.042 to -0.084, whereas a buyer-power story would predict attenuation once persistent buyer heterogeneity is absorbed. As a second check, we proxy buyer power by the number of distinct suppliers a buyer transacts with and interact $\ln q_{igjt}$ with this proxy. The interaction is significant at an economically negligible magnitude; full results are in Appendix B.3. Taken together, the evidence points to seller-side nonlinear pricing rather than buyer bargaining power as the primary driver of the observed patterns.

Nonlinear quantity discounts. Because products trade at different scales, we compare prices across ranks in each product's quantity distribution rather than raw quantities. For each product g, let $F_g(\cdot)$ be the empirical CDF of transacted quantities q_{igjt} using all 2024 observations of product g, and define the within-product rank:

$$r_{igjt} \;\equiv\; F_g\Big(q_{igjt}\Big)$$

We build a partition $[0,1]$ of 50 equal-probability intervals $I_b \equiv \Big((b-1)/50,\ b/50\Big]$ for $b = 1,\ldots,50$, and we assign each transaction to a bin $B_{igjt} = b$ whenever $r_{igjt} \in I_b$.[13] We then estimate

$$\ln p_{igjt} \;=\; \beta_0 \;+\; \sum_{b=2}^{50} \beta_b \,\mathbb{1}\Big\{B_{igjt} = b\Big\} \;+\; \Psi_{igd} \;+\; \varepsilon_{igjt} \tag{6}$$

[13]With discrete quantities and mass points, we assign observations to the smallest b such that $r_{igjt} \in I_b$;. Products with fewer than 50 distinct ranks are handled by the empirical CDF.

where Ψ_{igd} are seller-product-day fixed effects and bin b=1 (smallest-quantity bin) is the omitted category. By construction, bin b represents the same position in each product's quantity distribution, making the schedule comparable across heterogeneous products traded in heterogeneous units while absorbing all (i, g, d) shocks. Thus, identification of the coefficient comes solely from within-seller-product-day price variation.

Figure 3 summarizes the estimated schedule. Panel A plots the coefficients $\{\beta_b\}$ from (6) (with bin b=1 omitted) and, for readability, reports discounts relative to the smallest-quantity bin as $\Delta_b \equiv 1 - \exp(\beta_b)$. Prices fall steeply over the lower ranks: by b=10, unit prices are about 15% lower than in b=1. Discounts then continue to deepen but at a slower rate, stabilizing at around 17% for mid-to-large purchases. The final bin shows an additional dip, consistent with products or relationships concentrated in bulk trades.

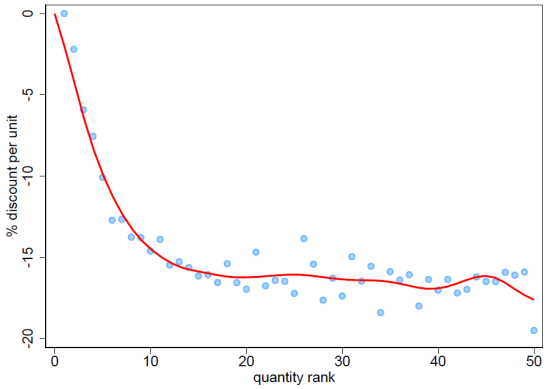
Price discrimination through a two-part tariff with flat fee F and constant marginal price p, as in Lemma 1, yields the average unit price $\bar{p}(q) = p + F/q$. Hence unit prices fall steeply at small q and flatten as q grows, approaching the constant marginal price p from above. The empirical schedule—steep discounts at low quantities and flattening at higher quantities—matches the prediction of Lemma 1, providing evidence consistent with a two-part tariff as the optimal nonlinear pricing strategy.

Figure 3 Panel B reports results from the rank-binning procedure that places each transaction by its position within that product's own quantity distribution, rather than by raw units. This lets us combine products that sell on very different scales while absorbing seller, buyer-group, and date effects. Because many products have fewer than 50 distinct quantity levels, bins are not uniform: common quantity "mass points" (e.g., single units or standard bulk packs) pile up at the extremes and some intermediate bins are empty. When ties occur, we assign the entire tied mass to the lower (earlier) bin, which mechanically pushes more observations into low-quantity bins. Consequently, about 55% of transactions fall into the first ten bins, where quantity discounts move the most; beyond roughly the 15th bin, the discount curve is comparatively flat.

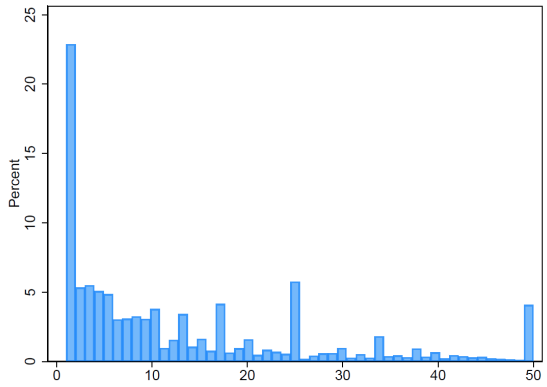
Heterogeneity across seller industries. We reestimate Equation (6) separately by 1-digit seller industry to compare price schedules across sectors. Figure 4 shows pronounced between-sector heterogeneity in both steepness and curvature. The Business Services and Construction sectors exhibit the largest declines—cumulative discounts approaching 30% by the top ranks—while the Manufacturing and Retail and Wholesale sectors display moderate but clear discounts (roughly 15%). The Transport and ICTs sector is comparatively flat, and the Financial Services sector is nearly flat across the entire rank distribution. Despite level differences, the qualitative shape (steep at low ranks, flattening at high ranks) is common, consistent with two-part tariffs where the fixed component is more salient for small purchases. In the model we develop in Section 4, this between-sector heterogeneity arises endogenously from differences in industry competition and in sellers'

Figure 3: Prices by Quantity Quantiles

Panel A. Quantity Discounts



Panel B. Histogram of Quantiles



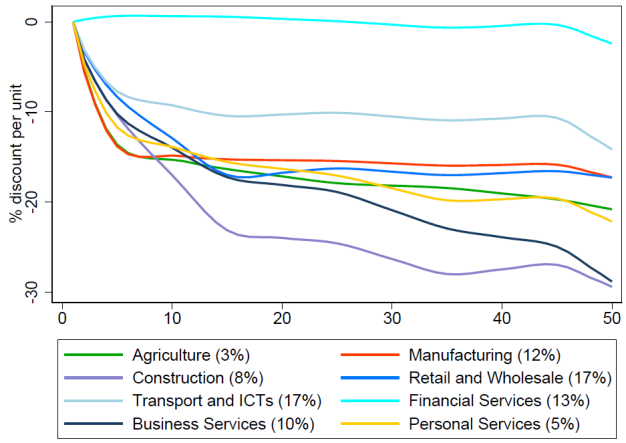
Notes: This figure summarizes the nonlinear relationship between quantity and price. Panel A plots the estimated coefficients from Regression (6), where log unit prices are regressed on 50 product-level quantity quantile indicators, controlling for seller-product-day fixed effects. The red line represents a fitted local polynomial of degree 5 of splines approximations of the estimated fixed effects. Panel B shows the distribution of observations across the modified quantiles, illustrating the rebinning strategy where each unique quantity is consistently mapped to a quantile across products. The first and last bins are overrepresented due to mass points in single and bulk purchases.

ability to appropriate buyer surplus.

Buyer-industry heterogeneity within seller sectors. We next fix a seller industry and reestimate Equation (6) separately by 1-digit buyer industry, recovering buyer-sector-specific schedules within each seller sector. Figure 5 illustrates the largest seller sectors by number of products transacted in 2024, Retail and Wholesale. Buyers in the Manufacturing, Retail and Wholesale, and Personal Services sectors exhibit sizable declines, whereas other buyer sectors display near-flat profiles (at most ≈5%, even at the top ranks). Because all specifications include seller-product-day fixed effects, these patterns reflect differential within-day, within-product price-rank relationships by buyer type. The evidence is consistent with a hybrid of second- and third-degree price discrimination: sellers deploy nonlinear schedules but tailor their menus to observable buyer characteristics. Mapping these results to Lemma 1, the evidence is consistent with heterogeneity in flat fees or marginal prices across buyer groups (i.e., from third-degree price discrimination), which shift the curves vertically.

Taking stock. Within seller-product-day cells, unit prices decline with quantity ranks and flatten at higher ranks. This curvature is pervasive across seller industries and shifts systematically with buyer industries (Figures 4 and 5). Because all (i, g, d) shocks are absorbed, these patterns reflect within-seller-product-day price differences inconsistent with uniform pricing and are consistent

Figure 4: Prices by Quantity Quantiles, by Seller Industry



Notes: This figure plots quantity discount schedules estimated separately by 1-digit seller industry. Each line represents a fifth-degree polynomial fit to splines approximations of the 50 fixed effects estimated from Equation (6), where the dependent variable is log unit price and the main regressor is a quantile bin of quantity, with seller-product-day fixed effects included. The y-axis measures the percentage discount per unit relative to the lowest-quantity transactions. The x-axis denotes the quantity quantile bin, ranging from 1 (smallest purchases) to 50 (largest). Sector labels show each industry's share of total GDP (excluding exports) in parentheses.

with a hybrid of second- and third-degree price discrimination: second-degree screening drives curvature, while observable buyer type shifts levels and steepness across industries. Guided by these facts, in the next section we develop a multisector supply-chain general equilibrium model with heterogeneous firms where endogenous contracts feature second- and third-degree price discrimination.

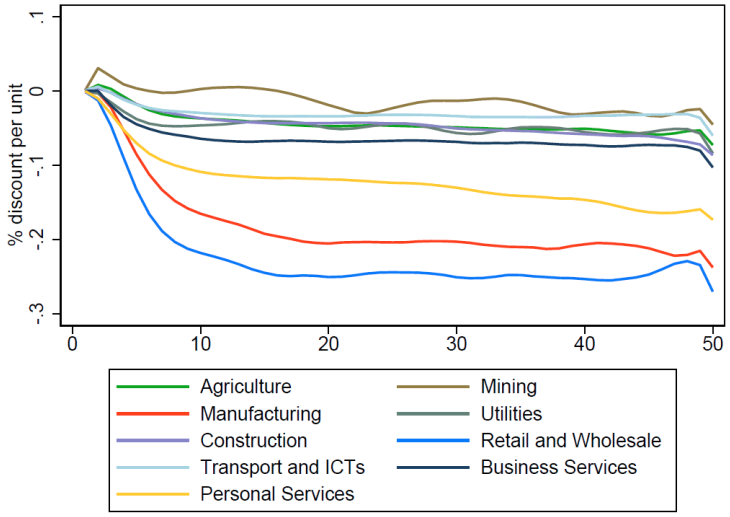
4 A Model of Nonlinear Pricing in Supply Chains

We develop and characterize a general-equilibrium supply-chain model in which firms simultaneously pay and charge nonlinear prices endogenously. We show that the optimal contract takes the form of a two-part tariff; a constant marginal price combined with a flat fee, extending Lemma 1 a general-equilibrium setting in supply chains. This framework provides closed-form sufficient statistics for welfare analysis under different pricing assumptions and counterfactual policies.

4.1 Environment and Notation

There are two firm types, $\ell \in \{u, r\}$, defined by their position with respect to final demand. Upstream firms (type u) sell to retailers and other upstream firms, and source inputs from upstream suppliers. Retailers (type r) purchase inputs from upstream firms and sell exclusively to the rep-

Figure 5: Prices by Quantity Quantiles: Retail and Wholesale Sellers, by Buyer Industry



Notes: We fix the seller industry and estimate Equation (6) separately for each 1-digit buyer industry. The fitted lines represent a fifth-degree polynomial fit to splines approximations of the 50 fixed effects estimated. Each curve corresponds to a specific buyer sector and traces the percentage discount per unit relative to the smallest purchases. The x-axis represents quantity quantiles from 1 (smallest) to 50 (largest).

resentative final consumer.[14] There is a finite set of sectors $\mathcal{S}$, common to both firm types. We use $s \in \mathcal{S}$ for buyer sectors and $s' \in \mathcal{S}$ for seller sectors. In each (ℓ, s), there is a continuum of firms that differ in their productivity z. Productivity z is Pareto-distributed within (ℓ, s) with lower bound $\underline{z}_s^\ell > 0$ and tail parameter $\kappa_s^\ell > 0$; the support is $z \in [\underline{z}_s^\ell, \infty)$.

When a firm appears as a buyer, we index it by i; when it appears as a seller, we index it by j. The mass of firms in (ℓ, s) is N_s^ℓ. We treat N_s^ℓ as endogenous under free entry; details follow. We evaluate the economy in steady state and omit time subscripts for brevity.

Market structure. Retailers sell to the representative consumer at uniform per-unit prices[15] while sourcing inputs from upstream firms at nonlinear prices. Upstream firms likewise purchase inputs from other upstream firms at nonlinear prices and sell their own variety to both retailers and other upstream firms. Consistent with the evidence we find for Chile, sellers j observe the buyer's type and sector pair (ℓ, s) but not the idiosyncratic buyer productivity z_i; they know only its (Pareto) distribution. They set type- and sector-specific tariff schedules but cannot condition on z_i, implying third-degree price discrimination across (ℓ, s) and second-degree price discrimination within

[14] This two-type structure is motivated by Chilean administrative data showing that most firms sell either only to final consumers or only to other firms, with minimal overlap; see Appendix B.4.

[15] For welfare effects of nonlinear pricing on final demand, see Bornstein and Peter (2024).

each (ℓ, s).[16]

Preferences. The representative consumer owns all firms and inelastically supplies one unit of labor (L=1). Let P_Y be the final-goods price index.[17] Final demand is Cobb-Douglas across retail sectors with a within-sector CES aggregator over retail varieties:

$$Y = \prod_{s \in \mathcal{S}} Y_s^{\theta_s}, \qquad \sum_{s \in \mathcal{S}} \theta_s = 1, \tag{7}$$

$$Y_s = \left(\int_{j \in \mathcal{R}_s} y_j^{\frac{\varphi_s - 1}{\varphi_s}} \, d\nu_s(j) \right)^{\frac{\varphi_s}{\varphi_s - 1}} \tag{8}$$

where $\theta_s \in (0, 1)$ are Cobb-Douglas output elasticities, $\varphi_s > 1$ is the within-sector elasticity of substitution, and $\mathcal{R}_s$ is the set of active retail sellers in sector s. Here, $d\nu_s(j)$ denotes the equilibrium measure over active retail sellers $j \in \mathcal{R}_s$, with total mass $N_s^r \equiv \nu_s(\mathcal{R}_s)$.

Technology. Firms (buyers i) produce with Cobb-Douglas technology in labor and a Cobb-Douglas aggregator across seller sectors; we denote the buyer sector by s and the seller sector by s':

$$Q_i = z_i \, l_i^{\alpha_s^\ell} \, M_i^{1 - \alpha_s^\ell}, \qquad 0 < \alpha_s^\ell < 1 \tag{9}$$

$$M_i = \prod_{s' \in \mathcal{S}} M_{is'}^{\theta_{ss'}^\ell}, \qquad \sum_{s' \in \mathcal{S}} \theta_{ss'}^\ell = 1 \quad \text{for each } (\ell, s) \tag{10}$$

where Q_i is output, z_i is firm i's productivity, l_i is firm-level labor input, α_s^ℓ is the labor output elasticity for firms of type ℓ in sector s, and M_i is the composite materials bundle. $M_{is'}$ is the materials bundle from upstream seller sector s', and $\theta_{ss'}^\ell \geq 0$ are input elasticities for buyers in (ℓ, s) across seller sectors $s' \in \mathcal{S}$.[18] Within any seller sector s', the materials bundle is CES across firm varieties with elasticity $\sigma_{s'} > 1$ for each $s' \in \mathcal{S}$:

$$M_{is'} = \left(\int_{j \in \mathcal{U}_{s'}} m_{ij}^{\frac{\sigma_{s'} - 1}{\sigma_{s'}}} \, d\nu_{s'}(j) \right)^{\frac{\sigma_{s'}}{\sigma_{s'} - 1}} \tag{11}$$

where m_{ij} is buyer i's input of seller variety j in seller sector s', $\sigma_{s'}$ is the elasticity of substitution across those varieties, and $d\nu_{s'}(j)$ is the equilibrium measure over active upstream sellers $\mathcal{U}_{s'}$, with total mass $N_{s'}^u \equiv \nu_{s'}(\mathcal{U}_{s'})$.

[16] Sustaining type- and sector-specific tariff schedules requires the absence of zero-cost arbitrage or resale; secondary markets may fail to emerge due to repackaging costs, regulation, or other frictions.

[17] If we normalize $P_Y \equiv 1$, welfare equals real final expenditure Y.

[18] A zero weight $\theta_{ss'}^\ell = 0$ means sector s as a buyer does not use inputs from sector s'. Under no price discrimination and uniform prices, $\{\theta_{ss'}^\ell\}$ coincide with input-output expenditure shares for buyer sector s, as in Acemoglu et al. (2012).

Input price-taking. Firms are atomistic in input markets and take the wage as given. They retain market power in output markets due to product differentiation under CES demand.

Firm entry. Firm entry follows Hopenhayn (1992) and Melitz (2003), adapted to a supply-chain environment. In each (ℓ, s), there is an unbounded pool of identical potential entrants. Entry requires paying a sunk cost $c_s^{E\,\ell} > 0$ in units of labor, after which firms draw their productivity z. Active firms exit exogenously at the end of the period with probability $\delta_s^\ell \in (0, 1]$, which serves as the only source of time discounting.[19] Let $\pi^{\ell s}(z)$ denote a potential entrant's per-period profit in numeraire units. Free entry requires that the expected discounted value of profits equals the entry cost in every (ℓ, s):

$$\frac{1}{1-\delta_s^\ell}\,\mathbb{E}_z\left[\pi^{\ell s}(z)\right] \;=\; c_s^{E\,\ell}\,w, \quad \forall(\ell, s)$$

where the expectation is taken over the postentry distribution of z in (ℓ, s).

Market clearing. All markets clear in equilibrium. Labor-market clearing requires that the total demand for labor across all active firms equals the inelastic supply of one unit:

$$\sum_{\ell\in\{u,r\}}\sum_{s\in\mathcal{S}}\int_{i\in\mathcal{F}_{\ell s}} l_i\,d\nu_{\ell s}(i) \;=\; L \;=\; 1$$

where $\mathcal{F}_{\ell s}$ is the set of active firms of type ℓ in sector s, and $\nu_{\ell s}$ is the equilibrium measure over these firms. For each upstream variety $j \in \mathcal{U}_{s'}$, market clearing requires that output equals the sum of inputs demanded by all buyers. For each retail variety $j \in \mathcal{R}_j$, market clearing requires that output equals final demand from the representative consumer:

$$Q_j \;=\; \sum_{\ell\in\{u,r\}}\sum_{s\in\mathcal{S}}\int_{i\in\mathcal{F}_{\ell s}} m_{ij}\,d\nu_{\ell s}(i), \quad \forall j \in \mathcal{U}_{s'},\ s' \in \mathcal{S}; \qquad Q_j \;=\; y_j \quad \forall s \in \mathcal{S},\ \forall j \in \mathcal{R}_s$$

General equilibrium under nonlinear pricing. Within a period, (i) potential entrants in each (ℓ, s) pay $c_s^{E\,\ell}$ and then draw productivity z; (ii) each upstream seller $j \in \mathcal{U}_{s'}$ observes only the buyer's pair (ℓ, s) (not z_i) and offers a pair-specific nonlinear contract menu $\{m_j^{\ell,s},\ T_j^{\ell,s}\}$; retail sellers $j \in \mathcal{R}_s$ post uniform prices to final consumers; (iii) buyers $i = (\ell, s, z_i)$ observe the offered menus and the wage w and choose labor l_i and input bundles $\{m_{ij}\}_j$ to maximize profits; (iv) production and trade occur, transfers $\{T_{ij}\}_j$ are realized, and final demand $\{y_j\}$ is met; and (v) firms

[19] Because we focus on steady-state comparisons of macroeconomic outcomes and abstract from time discounting aside from δ, the model is isomorphic to either a constant z over time or a stochastic process for z under the counterfactual of interest.

exit with probability δ_s^ℓ. Contracts are enforceable, resale and arbitrage are ruled out, and beliefs are rational; we consider a steady state so all aggregates are time-invariant.

A general equilibrium consists of allocations $\{Q_i,\, l_i,\, \{m_{ij}\}_j\}$, transfers $\{T_{ij}\}_j$, and consumer demands $\{y_j\}$ for all buyers $i = (\ell, s, z_i)$ with $\ell \in \{u, r\}$ and $s \in \mathcal{S}$, such that (i) technologies (9)–(11) hold for every firm; (ii) each upstream seller $j \in \mathcal{U}_{s'}$ chooses contracts $\{m_{ij}, T_{ij}\}_i$ that solve its profit-maximization problem (defined below), while each retail seller $j \in \mathcal{R}_s$ sets uniform prices to the final consumer; (iii) each buyer chooses labor l_i and input bundles $\{m_{ij}\}_j$ to maximize profits given the wage w and the offered contracts or prices; (iv) retail market clearing: $Q_j = y_j$ for all $s \in \mathcal{S}$ and all $j \in \mathcal{R}_s$; (v) upstream market clearing: $Q_j = \sum_{\ell \in \{u,r\}} \sum_{s \in \mathcal{S}} \int_{i \in \mathcal{F}_{\ell s}} m_{ij}\, d\nu_{\ell s}(i)$ for all $s' \in \mathcal{S}$ and all $j \in \mathcal{U}_{s'}$; (vi) the labor market clears; and (vii) free entry holds in each (ℓ, s).[20] A proof of existence and uniqueness is provided in Appendix C.7.

4.2 Guesses: Contracts and Revenue Shapes

We characterize the equilibrium using a guess-and-verify approach. In a supply-chain setting, firm costs and revenues are shaped by price discrimination. We begin by positing functional forms for contracts that deliver a tractable marginal cost function, which allows us to analyze the firm's unrestricted price-discrimination problem. We then verify the conjecture by showing that the resulting equilibrium coefficients are internally consistent. Motivated by Lemma 1, we conjecture that optimal contracts are equivalent to a two-part tariff specific to (ℓ, s). Furthermore, we conjecture that revenue functions are homogeneous of degree ψ_s^ℓ in output.

Guess 1: Two-part tariffs by buyer type and sector (ℓ, s). For a buyer $i = (\ell, s, z_i)$, the total payment for purchasing seller variety $j \in \mathcal{U}_{s'}$ is conjectured to take the form of a two–part tariff, with the marginal price determined by an (ℓ, s)-specific markup $\mu_{ss'}^\ell$:

$$T_{ij} \;=\; p_{js}^\ell\, m_{ij} \;+\; F_{js'}^\ell \;=\; \mu_{ss'}^\ell\, c_j\, m_{ij} \;+\; F_{js'}^\ell$$

where m_{ij} is the quantity purchased by buyer i from seller j, $p_{js}^\ell = \mu_{ss'}^\ell\, c_j$ is the marginal (allocative) price, ℓ and s denote the buyer's type and sector, and c_j is the seller's marginal cost. The fixed component, flat fee $F_{js'}^\ell$ varies with the seller identity j and with the buyer only through the observable pair (ℓ, s').

[20] Policy counterfactual equilibrium is defined analogously, except that firm maximization problems are subject to the additional constraints implied by the policy experiment (e.g., a ban on price discrimination or the introduction of output subsidies).

Guess 2: Equilibrium buyer revenue function. We conjecture that, in equilibrium, the revenue function is homogeneous of degree ψ_s^ℓ in output:

$$R_i \;=\; A_s^\ell \left(Q_i\right)^{\psi_s^\ell}$$

for unknown coefficients A_s^ℓ and ψ_s^ℓ that are constant at the buyer's type-sector (ℓ, s) level.

4.3 Preliminaries

To proceed, it is useful to derive input demand, price indices, and cost functions under the conjectured contract structure. Since flat fees are inframarginal, they do not affect these objects but only redistribute profits across firms. As a result, input demand, price indices, and cost functions coincide with those in a uniform-pricing economy, except that prices vary at the (ℓ, s) level. We then solve the unrestricted price-discrimination problem for a generic seller and verify the conjecture by matching undetermined coefficients.

4.3.1 Costs and Price Indices

Using the guesses in Section 4.2, in particular, that marginal prices are quantity-invariant within a buyer type–sector (ℓ, s) and a seller sector s', we can define sectoral price indices and derive firm costs. These objects will be verified once we solve for equilibrium prices.

CES sectoral price index. For any seller sector $s' \in \mathcal{S}$ and buyer type–sector (ℓ, s), let p_{js}^ℓ denote the marginal price charged by seller variety $j \in \mathcal{U}_{s'}$ to buyers in (ℓ, s). With elasticity $\sigma_{s'} > 1$, the unit price of the s'-bundle faced by buyers in (ℓ, s) is

$$P_{ss'}^\ell \;=\; \left(\int_{j \in \mathcal{U}_{s'}} \left(p_{js}^\ell\right)^{1-\sigma_{s'}} \, dv_{s'}(j)\right)^{\frac{1}{1-\sigma_{s'}}} \tag{12}$$

where $dv_{s'}(j)$ is the equilibrium measure over sellers in s', and $N_{s'}^u \equiv v_{s'}(\mathcal{U}_{s'})$ denotes their total mass. Flat fees do not enter (12).

Cobb-Douglas materials cost index. For firm $i = (\ell, s, z_i)$, the unit price of its composite materials bundle M_i in (10) is

$$P_i^M \;=\; \prod_{s' \in \mathcal{S}} \left(P_{ss'}^\ell\right)^{\theta_{ss'}^\ell}, \qquad \sum_{s' \in \mathcal{S}} \theta_{ss'}^\ell = 1, \;\; \theta_{ss'}^\ell \geq 0 \tag{13}$$

Firm-level marginal cost. Only marginal prices $\{p_{js}^\ell\}$ enter via (12) and (13); transfers T_{ij} are inframarginal and do not affect marginal cost. Given technology, wage $w > 0$, and constant returns

to scale, the marginal cost of producing Q_i units for firm $i = (\ell, s, z_i)$ is

$$c_i \;=\; \frac{\Theta_s^\ell}{z_i}\, w^{\alpha_s^\ell} \left(P_i^M\right)^{1-\alpha_s^\ell}, \qquad \text{where} \quad \Theta_s^\ell \;\equiv\; \left(\alpha_s^\ell\right)^{-\alpha_s^\ell} \left(1-\alpha_s^\ell\right)^{-(1-\alpha_s^\ell)} \prod_{s' \in \mathcal{S}} \left(\theta_{ss'}^\ell\right)^{-(1-\alpha_s^\ell)\theta_{ss'}^\ell}$$

Sectoral productivity index. Following Melitz (2003), define the CES sectoral productivity index for upstream (seller) sector s' and retail (seller) sector s as

$$\widetilde{z}_{s'}^{\,u} \;=\; \left(\int_{j \in \mathcal{U}_{s'}} z_j^{\sigma_{s'}-1}\, \frac{d\nu_{s'}(j)}{N_{s'}^u}\right)^{\frac{1}{\sigma_{s'}-1}}, \qquad \widetilde{z}_s^{\,r} \;=\; \left(\int_{j \in \mathcal{R}_s} z_j^{\varphi_s-1}\, \frac{d\nu_s(j)}{N_s^r}\right)^{\frac{1}{\varphi_s-1}}$$

4.3.2 Buyer Input Demand

Given the guesses in Section 4.2 and the objects defined in Section 4.3.1, each buyer $i = (\ell, s, z_i)$ chooses labor and input quantities from upstream seller varieties to maximize profits. Flat fees are inframarginal and do not affect marginal conditions; only marginal prices p_{is} matter for input choices. For notational simplicity, we therefore formulate the maximization problem in terms of profits $\widetilde{\Pi}_i$, net of flat fees.

Let m_{ij} denote the quantity of seller variety $j \in \mathcal{U}_{s'}$ purchased from seller sector s'. Using Guess 2, buyer $i = (\ell, s, z_i)$ solves

$$\widetilde{\Pi}_i = \max_{l_i,\, \{m_{ij}\}_j} \left\{ A_s^\ell\, Q_i^{\psi_s^\ell} \;-\; w\, l_i \;-\; \sum_{s' \in \mathcal{S}} \int_{j \in \mathcal{U}_{s'}} p_{js}^\ell\, m_{ij}\, d\nu_{s'}(j) \right\}$$

The total expenditure on inputs from seller sector s' can be expressed as $P_{ss'}^\ell\, M_{is'}$. The first-order condition with respect to $M_{is'}$ equates the marginal revenue product of the materials bundle to its sectoral price index, similarly for labor:

$$\frac{\partial R_i(z_i, \{M_{is'}\}, l_i)}{\partial M_{is'}} \;=\; P_{ss'}^\ell, \qquad \frac{\partial R_i(z_i, \{M_{is'}\}, l_i)}{\partial l_i} \;=\; w$$

which determines the labor-materials ratio given $\{P_{ss'}^\ell\}$. This input demand implies that the marginal revenue product of the materials bundle from sector s' is equalized across firm varieties within (ℓ, s). For a given buyer i, demand for the materials bundle from upstream sector s', denoted as $M_{is'}$, is allocated across varieties $j \in \mathcal{U}_{s'}$ according to the CES share rule. We have that p_{js}^ℓ is the price charged to buyers in (ℓ, s) and $P_{ss'}^\ell$ is the sectoral price index in (12). Under the conjecture that markups are (ℓ, s)-specific, it implies that relative input demands across varieties depend only on a seller's marginal cost relative to the sectoral price index. Buyer identity enters solely through

the scale term $M_{is'}$:

$$m_{ij} \;=\; M_{is'}\left(\frac{p^{\ell}_{js}}{P^{\ell}_{ss'}}\right)^{-\sigma_{s'}} \;=\; M_{is'}\left(\frac{\widetilde{z}^{u}_{s'}}{z_j\,(N^{u}_{s'})^{\frac{1}{1-\sigma_{s'}}}}\right)^{-\sigma_{s'}}, \qquad \sigma_{s'} > 1$$

where $\widetilde{z}^{u}_{s'}$ denotes the productivity index for sector s' and $N^{u}_{s'}$ is the measure of active upstream sellers in s'. Using this condition, together with market clearing, we can express total production of variety j as a function of its relative productivity:

$$Q_j \;=\; \left(\frac{z_j}{\widetilde{z}^{u}_{s'}}\right)^{\sigma_{s'}} Q_{s'}\big(\widetilde{z}^{u}_{s'}\big) \tag{14}$$

where $Q_{s'}(\widetilde{z}^{u}_{s'})$ denotes the total production of the average firm in upstream sector s'.

Scaling of input demand with productivity. Input usage scales with firm productivity relative to the sectoral average.[21] For an upstream firm $j \in \mathcal{U}_{s'}$ with productivity z_j, labor and material demand satisfy

$$l_j(z_j) \;=\; \left(\frac{z_j}{\widetilde{z}^{u}_{s'}}\right)^{\sigma_{s'}-1} l_j\big(\widetilde{z}^{u}_{s'}\big), \qquad M_j(z_j) \;=\; \left(\frac{z_j}{\widetilde{z}^{u}_{s'}}\right)^{\sigma_{s'}-1} M_j\big(\widetilde{z}^{u}_{s'}\big)$$

Hence, if productivities are Pareto-distributed with tail parameter $\kappa_{s'}$, input demand is also Pareto-distributed with a different tail parameter for upstream firms and retailers based on their relevant elasticity of substitution[22]:

$$\xi^{u}_{s'} \;=\; \frac{\kappa^{u}_{s'}}{\sigma_{s'}-1}, \qquad \xi^{r}_{s} \;=\; \frac{\kappa^{r}_{s}}{\varphi_s - 1} \;>\; 1$$

4.4 The Optimal Nonlinear Price

Fix an upstream seller sector s' and a buyer type–sector (ℓ, s). A seller $j \in \mathcal{U}_{s'}$ offers an unrestricted menu $\{x, T\}$ to buyers $i = (\ell, s, z_i)$, where x denotes the allocated quantity and T the transfer.

To describe the buyer i surplus and the extractable rents by seller j when transacting with i, let $\nu_{s'}$ denote the equilibrium measure over upstream sellers in sector s'. Denote buyer i's profit (inclusive of transfers) by Π_i. For buyer i, the total surplus from transacting with seller j of pro-

[21] This follows from the homogeneity of the Cobb-Douglas technology: $y(x_1, \ldots) = x_1 \cdot y(1, x_2/x_1, \ldots)$ implies that input ratios are pinned down by common input prices. Scaling by relative productivity yields $\frac{l(z)}{l(\bar{z})} = (\frac{z}{\bar{z}})^{\sigma-1}$, so input demand inherits a Pareto distribution with effective tail parameter $\xi = \kappa/(\sigma-1)$.

[22] Equilibrium feasibility requires that $\xi^{u}_{s'} > 1$ and $\xi^{r}_{s} > 1$. Aggregate labor demand is $L = \int l(z)\, d\nu(z)$, which is finite only if $\xi > 1$; hence, $\kappa^{u}_{s'} > \sigma_{s'} - 1$ upstream and $\kappa^{r}_{s} > \varphi_s - 1$ in retail.

ductivity z_j is defined as

$$TS_{is'}(m_{ij}) := \left.\frac{d\,\Pi_i}{d\big(\nu_{s'}(z_j)\big)}\right|_{\arg\max \Pi_i}$$

namely, the marginal value to buyer i of access to an additional infinitesimal mass of sellers of type z_j within sector s', evaluated at buyer i optimal input choices. This is the surplus that an infinitesimal seller j seeks to appropriate through its contract. We can express this as surplus in terms of the marginal revenue product. Under CES aggregation within sector u' (elasticity $\sigma_{u'} > 1$), the extractable surplus from a purchase of a generic size m can be written as

$$TS_{is'}(m) \;=\; \frac{\sigma_{s'}}{\sigma_{s'}-1}\,\frac{\partial R_i(z_i,\{M_{is'}\},l_i)}{\partial M_{is'}}\,M_{is'}^{\frac{1}{\sigma_{s'}}}\;m^{\frac{\sigma_{s'}-1}{\sigma_{s'}}} \;-\; T$$

where m is the quantity purchased and T the associated transfer. Since the individual seller is infinitesimal, it treats the marginal revenue product as given. Using the first-order condition, this simplifies to

$$TS_{is'}(m_{ij}) \;=\; \frac{\sigma_{s'}}{\sigma_{s'}-1}\,P_{js'}^{\ell}\,M_{is'}^{\frac{1}{\sigma_{s'}}}\;m^{\frac{\sigma_{s'}-1}{\sigma_{s'}}} \;-\; T$$

Valuation index and its distribution. For a seller in sector s', the buyer i matters only through the one-dimensional valuation index:

$$\tau_{is'} \;\equiv\; P_{ss'}^{\ell}\,M_{is'}^{1/\sigma_{s'}}$$

The valuation index, $\tau_{is'}$, combines the sector s' price level $P_{ss'}^{\ell}$ with the buyer's scale $M_{is'}$ in the exact way that determines the marginal revenue from a small purchase: from the CES share rule, a seller's marginal revenue is proportional to $\tau_{is'}\,m^{(\sigma_{s'}-1)/\sigma_{s'}}$ for quantity m (up to the transfer T). Hence, a seller's problem can be written solely in terms of $\tau_{is'}$.

When buyer productivities z_i in (ℓ, s) are Pareto, $\tau_{is'}$ is Pareto as well. Let $M_{is'}$ be the buyer's materials demand from sector s'; under our technology, $M_{is'}$ is strictly increasing in z_i and is distributed according to a Pareto with tail ξ_s^{ℓ}. Therefore, $\tau_{is'}$ inherits the Pareto law of the productivity distribution z, with

$$\tau_{is'} \;\sim\; \text{Pareto}\big(\rho_{ss'}^{\ell}\big), \qquad \rho_{ss'}^{\ell} \;=\; \sigma_{s'}\,\xi_s^{\ell}$$

Rescaling by $P_{ss'}^{\ell}$ shifts only the scale (not the tail) of the distribution. Feasibility requires $\xi_s^{\ell} > 1$, which implies $\rho_{ss'}^{\ell} > \sigma_{s'}$ for all (ℓ, s, s').

Seller's problem. Applying the revelation principle, a seller in sector s' chooses menus of allocations $x(\tau)$ and transfers $T(\tau)$ for all buyers $i = (\ell, s, z_i)$. The total profit-maximization problem

is

$$\max_{\{x(\cdot),\,T(\cdot)\}} \sum_{\ell\in\{u,r\}} \sum_{s\in\mathcal{S}} N_s^{\ell}\, \mathbb{E}_{\tau_{is'}}\Big[\, T(\tau) - c_j\, x(\tau)\,\Big] \tag{15}$$

subject to

$$\text{(LIC)}\quad TS_{ss'}^{\ell}(\tau) = TS_{ss'}^{\ell}(\underline{\tau}) + \frac{\sigma_{s'}}{\sigma_{s'}-1}\int_{\underline{\tau}}^{\tau} x(\omega)^{\frac{\sigma_{s'}-1}{\sigma_{s'}}}\, d\omega,$$

$$\text{(IR)}\quad TS_{ss'}^{\ell}(\underline{\tau}) \;\geq\; 0,$$

$$\text{(Monotonicity)}\quad x(\tau') \;\geq\; x(\tau) \quad \text{for } \tau' > \tau$$

The local incentive-compatibility constraint (LIC) ensures that truth-telling is optimal for all buyer types τ locally; (IR) is the individual-rationality constraint, binding for the lowest type $\underline{\tau}$; and monotonicity requires that higher buyer types receive weakly larger allocations. Taken together, (LIC) and monotonicity guarantee that the mechanism is incentive-compatible globally. Because the objective and constraints are additively separable across (ℓ, s, s') triples, the problem can be solved independently for each partition.

Solution concept. The seller's mechanism design problem is equivalent to the setting of Lemma 1, and we solve it via the virtual-surplus approach. The sufficient conditions in Lemma 1 are satisfied: marginal cost is constant, the revenue function is homogeneous in output, and buyer types are Pareto-distributed. With one-dimensional type $\tau_{is'}$ and quasilinear transfers, expected revenue equals expected virtual surplus. Under the regularity condition $\rho_{ss'}^{\ell} > \sigma_{s'}$ (increasing virtual value), the problem separates across (ℓ, s) for a given seller sector s' and is solved pointwise in τ; transfers follow from the envelope formula with IR binding at $\underline{\tau}$.

Proposition 1 (Optimal Nonlinear Price for Upstream Sellers)**.** *In equilibrium, the optimal contract offered by an upstream seller $j \in \boldsymbol{\mathcal{U}}_{s'}$ to any buyer $i = (\ell, s, z_i)$ is a two-part tariff:*

$$T_{ij} \;=\; p_{js}^{\ell}\, m_{ij} \;+\; F_{js}^{\ell}$$

with marginal (allocative) price

$$p_{js}^{\ell} \;=\; \mu_{ss'}^{\ell}\, c_j, \qquad \mu_{ss'}^{\ell} \;=\; \frac{\rho_{ss'}^{\ell}}{\rho_{ss'}^{\ell}-1}, \qquad \rho_{ss'}^{\ell} \;=\; \xi_s^{\ell}\, \sigma_{s'}$$

a constant markup over marginal cost ***within each buyer type–sector*** (ℓ, s) *for a given seller sector s'. The fixed component F_{js}^{ℓ} is chosen so that the lowest buyer type obtains zero surplus:*

$$F_{js}^{\ell} \;=\; \frac{1}{\sigma_{s'}-1}\,\tau_{is'}(\underline{z})\left(M_{is'}(\underline{z})\right)^{\frac{\sigma_{s'}-1}{\sigma_{s'}}}\left(\frac{p_{js}^{\ell}}{P_{ss'}^{\ell}}\right)^{1-\sigma_{s'}} \;=\; \left(\frac{z_j}{\underline{z}_{s'}^{u}}\right)^{\sigma_{s'}-1}\overline{F}_{ss'}^{\ell}, \qquad \tau_{is'} \;=\; P_{ss'}^{\ell}\, M_{is'}^{1/\sigma_{s'}}$$

*Here, $\overline{F}^{\ell}_{ss'}$ denotes the **average flat fee per seller** in sector s':*

$$\overline{F}^{\ell}_{ss'} \;\equiv\; \frac{1}{N^{u}_{s'}}\,\frac{1}{\sigma_{s'}-1}\,\tau_{is'}(\underline{z})\left(M_{is'}(\underline{z})\right)^{\frac{\sigma_{s'}-1}{\sigma_{s'}}}$$

Proof sketch and verification of the guesses. For upstream sellers (virtual surplus), fix a seller sector s' and a seller $j \in \mathcal{U}_{s'}$. Because the objective and constraints are additively separable across buyer partitions (ℓ, s), the mechanism is solved partition-by-partition. By Lemma 1 with type $\tau_{is'}$ (Pareto tail $\rho^{\ell}_{ss'}$), the partition problem is

$$\max_{x(\tau)}\; N^{\ell}_{s}\,\mathbb{E}_{\tau_{is'}}\left[\left(\tau - g^{-1}(\tau)\right)\frac{\sigma_{s'}}{\sigma_{s'}-1}\,x(\tau)^{\frac{\sigma_{s'}-1}{\sigma_{s'}}} \;-\; c_j\,x(\tau)\right]$$

For Pareto, $g^{-1}(\tau) = \tau/\rho^{\ell}_{ss'}$, so the virtual value is increasing when $\rho^{\ell}_{ss'} > \sigma_{s'}$. The FOC yields a constant markup within the partition:

$$p^{\ell}_{js} \;=\; \mu^{\ell}_{ss'}\,c_j, \qquad \mu^{\ell}_{ss'} \;=\; \frac{\rho^{\ell}_{ss'}}{\rho^{\ell}_{ss'}-1}$$

and the fixed component is pinned down by IR at the lowest buyer type $\underline{z}^{\ell}_{s}$:

$$F^{\ell}_{js} \;=\; \frac{1}{\sigma_{s'}-1}\,\tau_{is'}(\underline{z}^{\ell}_{s})\left(M_{is'}(\underline{z}^{\ell}_{s})\right)^{\frac{\sigma_{s'}-1}{\sigma_{s'}}}\left(\frac{p^{\ell}_{js}}{P^{\ell}_{ss'}}\right)^{1-\sigma_{s'}}$$

For upstream sellers (homogeneous revenue), constant allocative prices $p^{\ell}_{js} = \mu^{\ell}_{ss'}c_j$ and the CES share rule imply, after aggregating over buyers and partitions, that total output Q_j is proportional to $c_j^{-\sigma_{s'}}$ times an aggregate demand term that depends on buyer masses, price indices, and materials bundles (see Appendix C.2). Equivalently, $c_j^{1-\sigma_{s'}}$ is proportional to $Q_j^{(\sigma_{s'}-1)/\sigma_{s'}}$. Both the per-unit revenue and the fee component scale with the same CES share, so total revenue scales as $R_j \;\propto\; Q_j^{(\sigma_{s'}-1)/\sigma_{s'}}$, up to a sector-$s'$ constant that aggregates buyer-side objects. This verifies Guess 1 (two-part tariffs) and Guess 2 (homogeneous revenues) for upstream sellers.

For retailers ($\ell = r$), uniform pricing under within-sector CES demand yields revenue proportional to $Q_j^{(\varphi_s-1)/\varphi_s}$ with a shifter depending on the retail price index P_s and expenditure $\theta_s Y$ (with $P_Y \equiv 1$), as shown in Appendix C.1. This completes the verification of Guess 2 for retailers and, together with the upstream case, confirms both guesses.

4.5 Other Pricing Regimes for Welfare Comparisons

We compare the nonlinear-pricing benchmark to two counterfactual policies: (i) monopolistic competition with uniform prices, which corresponds to a complete ban on price discrimination;

and (ii) a planner-implemented allocation in a decentralized equilibrium, which is attained with a ban on price discrimination and an output subsidy that restore marginal cost pricing conditional on entry.

Monopolistic competition (uniform pricing). Under uniform pricing, each upstream seller $j \in \mathcal{U}_{s'}$ charges the same CES markup over marginal cost to all buyers, regardless of their partition (ℓ, j):

$$p_{js}^{\ell,\text{Lin}} \;=\; \mu_{s'}^{\text{Lin}}\, c_j \qquad \mu_{s'}^{\text{Lin}} \;\equiv\; \frac{\sigma_{s'}}{\sigma_{s'}-1}$$

where c_j is the marginal cost of seller firm j. Thus, in contrast to nonlinear pricing, allocative prices do not vary across buyer partitions.

Retailers in sector s sell to final demand at the CES markup,

$$\mu_s^{r,\text{Lin}} \;=\; \frac{\varphi_s}{\varphi_s - 1}$$

For buyers of type ℓ in sector s, the CES price index for inputs from upstream sector q is

$$P_{ss'}^{\ell,\text{Lin}} = \left(\int_{j\in\mathcal{U}_{s'}} \left(p_{js}^{\ell,\text{Lin}}\right)^{1-\sigma_{s'}} dv_{s'}(j)\right)^{\frac{1}{1-\sigma_{s'}}} = \mu_{s'}^{\text{Lin}} \left(\int_{j\in\mathcal{U}_{s'}} c_j^{\,1-\sigma_{s'}}\, dv_{s'}(j)\right)^{\frac{1}{1-\sigma_{s'}}}$$

where $dv_{s'}(j)$ integrates over active upstream firms in sector s', with free entry in each (ℓ, s).

Lemma 2 (Efficiency with CES markups and per-unit output subsidies)**.** *Consider the economy under a complete ban on price discrimination, so all sellers post uniform prices: each upstream seller sector $s' \in \mathcal{S}$ and retail sector $s \in \mathcal{S}$ sets the CES markup over marginal cost. The efficient allocation is achieved if the government rebates a per-unit output subsidy that restores marginal-cost pricing conditional on entry:*

$$p_j^{Lin} \;=\; \mu_{s'}^{Lin}\, c_j, \qquad \mu_{s'}^{Lin} \;=\; \frac{\sigma_{s'}}{\sigma_{s'}-1}, \qquad \tau_{s'}^{u} \;=\; \left(1 - \frac{1}{\mu_{s'}^{Lin}}\right) c_j \;=\; \frac{1}{\sigma_{s'}}\, c_j;$$

$$p_i^{r,Lin} \;=\; \mu_s^{r,Lin}\, c_i, \qquad \mu_s^{r,Lin} \;=\; \frac{\varphi_s}{\varphi_s - 1}, \qquad \tau_s^{r} \;=\; \left(1 - \frac{1}{\mu_s^{r,Lin}}\right) c_i \;=\; \frac{1}{\varphi_s}\, c_i$$

Then the resulting decentralized equilibrium is efficient.

This lemma is a special case of the general result in Theorem 1 of Baqaee and Farhi (2020a) (details in Appendix C.3). Efficiency is obtained in a decentralized equilibrium when each variety charges a markup equal to its consumer-surplus ratio and receives output subsidies that exactly offset the induced within-period pricing wedge. In our CES setting, the consumer-surplus ratio for an upstream variety coincides with $\mu_{s'}^{\text{Lin}}$, and for a retail variety with $\mu_s^{r,\text{Lin}}$. Though charging the CES markup delivers the correct expected profits and thereby ensures efficient entry, it distorts

input choices by acting as a tax on production. An output subsidy is therefore required to undo this distortion and restore marginal-cost pricing conditional on entry.

4.6 Theoretical Results

We now collect the main equilibrium implications of nonlinear pricing in our supply-chain model.

Result 1: allocative markups under nonlinear pricing. For any buyer partition (ℓ, s) and seller sector s', the allocative markup under nonlinear pricing is strictly below the uniform-CES markup:

$$\mu_{ss'}^{\ell} \;=\; \frac{\rho_{ss'}^{\ell}}{\rho_{ss'}^{\ell}-1} \;<\; \frac{\sigma_{s'}}{\sigma_{s'}-1} \quad \text{since} \quad \rho_{ss'}^{\ell} \;=\; \xi_{s}^{\ell}\,\sigma_{s'} \text{ with } \xi_{s}^{\ell} > 1$$

Under uniform pricing, the markup is determined by the elasticity of substitution. In nonlinear pricing, it is instead determined by a combination of the elasticity of substitution and the Pareto-tail distribution parameter. This aligns marginal revenue more closely with the shape of demand, because price discrimination allows the seller to extract surplus through flat fees rather than distorting marginal allocations. Consequently, nonlinear pricing reduces allocative distortions at the margin relative to uniform pricing.

Result 2: Seller-identity invariance of the total unit markup. Fix a seller sector s' and a buyer partition (ℓ, s). For any seller $j \in \mathcal{U}_{s'}$ and buyer $i = (\ell, s, z_i)$, the per-unit payment decomposes as

$$\frac{T_{ij}}{m_{ij}} \;=\; p_{js}^{\ell} \;+\; \frac{F_{js}^{\ell}}{m_{ij}}$$

The resulting total unit markup (unit price over marginal cost) satisfies

$$\frac{\frac{T_{ij}}{m_{ij}}}{c_j} \;=\; \mu_{ss'}^{\ell}\left(1+\chi_{ss'}^{\ell}(i)\right), \qquad \mu_{ss'}^{\ell} = \frac{\rho_{ss'}^{\ell}}{\rho_{ss'}^{\ell}-1}$$

where $\chi_{ss'}^{\ell}(i)$ is a buyer-specific scalar defined in Appendix C.4. It depends on (ℓ, s) objects (the sectoral price index and the buyer's sector–s' bundle, including the lowest buyer type) but not on the seller j. Hence, within a given buyer partition, the total unit markup is invariant to the seller's identity.

Result 3: Average flat fee paid to seller sector s' determinants. For any buyer partition (ℓ, s) with lowest type $\underline{z}_s^\ell$, the average flat fee paid to seller sector s' is

$$\overline{F}_{ss'}^{\ell} \;=\; \frac{P_{ss'}^{\ell}\, M_{is'}(\underline{z}_s^{\ell})}{N_{s'}^{u}\,(\sigma_{s'}-1)} \;=\; \frac{\psi_s^{\ell}}{\sigma_{s'}-1}\,(1-\alpha_s^{\ell})\,\theta_{ss'}^{\ell}\,\frac{R_s^{\ell}(\underline{z}_s^{\ell})}{N_{s'}^{u}}$$

Here, $R_s^\ell(\underline{z}_s^\ell)$ denotes the revenue of the lowest–type buyer in (ℓ, s), $M_{is'}(\underline{z}_s^\ell)$ the corresponding sector-s' materials bundle, $P_{ss'}^\ell$ the sectoral price index faced by (ℓ, s), and $N_{s'}^u$ the mass of active upstream sellers in s'.

Five forces shape flat fees from buyers in (ℓ, s) to seller sector s': (i) lowest-type revenue $R_s^\ell(\underline{z}_s^\ell)$ raises extractable rents; (ii) revenue curvature ψ_s^ℓ (from $R_i = A_s^\ell Q_i^{\psi_s^\ell}$) scales marginal surplus—more concavity lowers fees; (iii) input importance $(1-\alpha_s^\ell)\theta_{ss'}^\ell$ increases the surplus a seller can extract; (iv) a larger $\sigma_{s'}$ (greater substitutability) increases competition, reducing the market power and thus the surplus a seller can extract through fees; and (v) a larger $N_{s'}^u$ dilutes rents across more sellers, lowering the average fee.

Result 4: Firm profits rely on flat fees. With two–part tariffs, profits decompose into a marginal (per-unit) component and a fixed (flat-fee) component. This holds for upstream sellers and for retailers. This decomposition highlights which forces move profits: allocative markups on the margin, and the incidence of fixed transfers across buyer-seller pairs. For an upstream seller $j \in \mathcal{U}_{s'}$,

$$\mathbb{E}\left[\Pi_j^u\right] = \underbrace{\sum_{\ell\in\{u,r\}}\sum_{s\in\mathcal{S}}\int_{i\in\mathcal{F}_{\ell s}}\left(p_{js}^{\ell}-c_j\right)m_{ij}\,d\nu_{\ell s}(i)}_{\text{allocative margin}} + \underbrace{\sum_{\ell\in\{u,r\}}\sum_{s\in\mathcal{S}}\int_{i\in\mathcal{F}_{\ell s}}F_{js}^{\ell}(i)\,d\nu_{\ell s}(i)}_{\text{flat-fee revenue}} - \underbrace{\sum_{t\in\mathcal{S}}\int_{h\in\mathcal{U}_t}F_{h\,s'}^{u}(j)\,d\nu_t(h)}_{\text{flat-fee payments}}$$

For a retailer $i \in \mathcal{R}_s$,

$$\mathbb{E}\left[\Pi_i^r\right] = \underbrace{\left(\tfrac{1}{\varphi_s}\right)R_i}_{\text{allocative margin}} - \underbrace{\sum_{s'\in\mathcal{S}}\int_{j\in\mathcal{U}_{s'}}F_{js}^{r}(i)\,d\nu_{s'}(j)}_{\text{fees to upstream}}$$

Realized profits depend on the specific network of trading partners (which sellers a buyer contracts with, and which upstream tiers a seller sources from). For sector-level analysis, we therefore work with expected profits; averages over the equilibrium measures $\nu_{\ell s}$ (buyers in partition (ℓ, s)) and $\nu_{s'}$ (sellers in sector s'), which collapse partner-specific parameters into sectoral aggregates.

Here, p_{js}^ℓ is the allocative (marginal) price charged by seller j to buyers in (ℓ, s), c_j is seller j's marginal cost, m_{ij} is buyer i's quantity purchased from j, and $F_{js}^\ell(i)$ is the flat fee paid by buyer i to seller j. The set $\mathcal{F}_{\ell s}$ collects active buyers in (ℓ, s) with measure $\nu_{\ell s}$; $\mathcal{U}_t$ is the set of upstream sellers

in sector t with measure ν_t. The payment $F^u_{hs'}(j)$ denotes the flat fee that seller j (as a buyer of type u in buyer sector s') pays to its upstream supplier $h \in \mathcal{U}_t$.

The allocative margin captures the usual markup-cost wedge times purchased quantities; under nonlinear pricing, Result 1 implies that these markups are lower than under uniform pricing, shrinking this component. The flat-fee terms redistribute surplus based on each seller's CES share in a buyer's materials bundle and on how important inputs are in production (via the θ's). In expectation, the network of bilateral contracts integrates out to sectoral objects, so expected upstream profits can be expressed as affine functions of sectoral labor expenditures (Appendix C.5), and profits of lowest-productivity retailers hinge on input substitutability and input cost shares (Appendix C.5). allocative markups.

The fact that profits need not vanish contrasts with the standard mechanism-design benchmark, in which the lowest type's surplus is pinned to zero. Here, bilateral surplus is zero at the margin of each input transaction, but integrating across all transactions leaves residual profits (through labor, which is not price-discriminated) or, conversely, negative profits when intermediates are insufficiently substitutable. In the Leontief limit, each supplier's marginal contribution equals the buyer's total surplus, so every supplier attempts to appropriate the full rent. This drives the lowest type's profit below zero and generates a hold-up problem that deters entry.

Flat fees are inframarginal: they do not affect first-order input choices or final demand, but they do reallocate surplus along the chain. With a representative owner, these transfers net out at a point in time; in general equilibrium, however, they shift free-entry conditions across (ℓ, s), altering the mass of active varieties and sectoral price indices. Hence, welfare in the counterfactuals will be shaped by two channels: (i) changes in allocative markups (marginal wedges), and (ii) the reallocation of flat-fee income that tilts entry across sectors. In the next section, we formalize these channels and map welfare changes to sufficient statistics tied to markups, price indices, and entry margins.

4.7 Welfare Decomposition: Intensive vs. Extensive Margins

In this section, we represent changes in welfare as a function of sectoral markups and sectoral firm masses. We set the wage as numeraire and under free entry, aggregate welfare is the inverse of the final-good price index, $W \equiv 1/P_Y$. The price index satisfies $\log P_Y = \sum_{s\in\mathcal{S}} \theta_s \log P_s$, where θ_s are final-expenditure shares.

To decompose welfare, we introduce input-output objects. We define the row vector of retail final-demand shares as $b := (\theta_s)_{s\in\mathcal{S}} \in \mathbb{R}^{1\times|\mathcal{S}|}$ and Ω as the cost-based input-output matrix stacking retail and upstream sectors, of dimensions $2|\mathcal{S}| \times 2|\mathcal{S}|$. Each element of Ω captures the direct cost exposure of buyers to their upstream input suppliers. We single out two $|\mathcal{S}| \times |\mathcal{S}|$ blocks:

$$\Omega^{uu} \in \mathbb{R}^{|\mathcal{S}|\times|\mathcal{S}|}, \qquad \Omega^{ru} \in \mathbb{R}^{|\mathcal{S}|\times|\mathcal{S}|}$$

with elements, for buyer sector s and seller sector s',

$$\Omega^{uu}_{ss'} := (1 - \alpha^u_s)\,\theta^u_{ss'}, \qquad \Omega^{ru}_{ss'} := (1 - \alpha^r_s)\,\theta^r_{ss'}$$

The upstream cost-based Leontief inverse, which accounts for direct and indirect exposures through the supply chain, is

$$\Psi^{uu} := (I - \Omega^{uu})^{-1}$$

Because the final-good aggregator is separable across retail sectors, the retail-consumer map is the identity, so b already reflects direct retail exposure. We then define sectoral final-demand exposures (in the style of Baqaee and Farhi (2020b)) as scalars, for each $s \in \mathcal{S}$:

$$\tilde{\lambda}^{cr}_s := b_s, \qquad \tilde{\lambda}^{ru}_s := \sum_{v\in\mathcal{S}} \tilde{\lambda}^{cr}_v\,\Omega^{ru}_{vs}, \qquad \tilde{\lambda}^{uu}_s := \sum_{v\in\mathcal{S}} \tilde{\lambda}^{ru}_v\,\Psi^{uu}_{vs}$$

and it, in vector form,

$$\tilde{\lambda}^{cr} = b, \qquad \tilde{\lambda}^{ru} = b\,\Omega^{ru}, \qquad \tilde{\lambda}^{uu} = b\,\Omega^{ru}\,\Psi^{uu}$$

Let μ^r_s denote the retail-consumer markup in sector s (stacking over s yields a vector in $\mathbb{R}^{|\mathcal{S}|}$). Let $\mu^r \in \mathbb{R}^{|\mathcal{S}|}$ collect the retail-upstream markups by sector, and let $\mu^{uu} \in \mathbb{R}^{|\mathcal{S}|}$ collect the upstream-upstream markups by sector. Firm masses are N^r_s for retail sector s and N^u_s for upstream sector s, stacked as vectors $N^r, N^u \in \mathbb{R}^{|\mathcal{S}|}$. Elasticities are $\varphi_s > 1$ (retail) and $\sigma_s > 1$ (upstream), stacked as $\varphi, \sigma \in \mathbb{R}^{|\mathcal{S}|}$.

Proposition 2 (Exact welfare decomposition[23])**.** *Fix the exposure scalars $\{\tilde{\lambda}^{cr}_s, \tilde{\lambda}^{ru}_s, \tilde{\lambda}^{uu}_s\}_{s\in\mathcal{S}}$ defined above and sectoral objects $\{\mu^r_s\}_{s\in\mathcal{S}}$, $\mu^r \in \mathbb{R}^{|\mathcal{S}|}$, $\mu^{uu} \in \mathbb{R}^{|\mathcal{S}|}$, $N^r, N^u \in \mathbb{R}^{|\mathcal{S}|}$, $\varphi, \sigma \in \mathbb{R}^{|\mathcal{S}|}$ with $\varphi_s, \sigma_s > 1$ The change in welfare satisfies*[24]

$$\Delta \log W = \underbrace{-\sum_{s\in\mathcal{S}} \tilde{\lambda}^{cr}_s\,\Delta\log\mu^r_s \;-\; \tilde{\lambda}^{ru}\,\Delta\log\mu^r \;-\; \tilde{\lambda}^{uu}\,\Delta\log\mu^{uu}}_{\textit{Intensive margin (markups)}} + \underbrace{\sum_{s\in\mathcal{S}} \frac{\tilde{\lambda}^{cr}_s}{\varphi_s - 1}\,\Delta\log N^r_s \;+\; \frac{\tilde{\lambda}^{uu}}{\sigma - 1}\Delta\log N^u}_{\textit{Extensive margin (firm masses)}}$$

The first brace (intensive margin) captures the allocative effect of markups on the final-good price index. Changes in retail-consumer markups are weighted by $\tilde{\lambda}^{cr}_s$; buyer-specific retail-upstream wedges load with the components of $\tilde{\lambda}^{ru}$; and upstream–upstream wedges load with the compo-

[23] This proof comes from bulding a linear system linking sectoral price index changes to markup shocks and changes in the mass of active firms via the input-ouput matrix Ω. Solving this system provides an exposure map from (i) markups to prices (intensive margin) and (ii) variety/entry to prices (extensive margin), both with final-demand weights. Full derivations are in Appendix C.6.

[24] Here, $\tilde{\lambda}^{ru}\,\Delta\log\mu^r$ and $\tilde{\lambda}^{uu}\,(\Delta\log N^u/(\sigma-1))$ denote row–column contractions; the division by $(\sigma - 1)$ is elementwise. All $\Delta\log(\cdot)$ act componentwise across sectors. If composition (selection) effects were present, an additional term involving CES selection objects would enter the identity; under our Pareto-type assumption, composition is invariant and that term is zero.

nents of $\tilde{\lambda}^{uu}$. Sectors with larger exposure to final demand therefore exert a disproportionately large influence on $\log W$: a given percentage change in a high-exposure sector's markup moves aggregate welfare more.

The second brace (extensive margin) captures how the masses of active varieties affect variety-adjusted price indices. Retail entry loads with $\tilde{\lambda}_s^{cr}/(\varphi_s - 1)$, while upstream entry loads with $\tilde{\lambda}^{uu} \odot (\sigma - 1)^{-1}$ (elementwise). The exposure vectors $\tilde{\lambda}^{cr}, \tilde{\lambda}^{ru}, \tilde{\lambda}^{uu}$ act as general-equilibrium multipliers—closely analogous to cost-based Domar weights in more general production-networkmultipliers models ((Baqaee and Farhi, 2020b))—that translate sectoral markups and entry responses into aggregate welfare changes. Hence, losses from market power and from variety distortions are weighted by position in the supply chain and exposure to final demand.

Comparing pricing regimes amounts to evaluating the same decomposition across different equilibria. Nonlinear pricing weakens allocative wedges relative to uniform pricing, unambiguously raising the intensive margin of welfare, with gains scaled by the exposure vectors. Flat fees are inframarginal and operate through participation and entry; the extensive component can therefore amplify or offset the intensive gains depending on how firm masses adjust in highly exposed sectors. Overall, the net welfare effect is governed by the exposure maps and the induced general-equilibrium entry responses; we quantify these forces in the next section.

5 Model Calibration and Quantification

Using population-level data on firm-to-firm transactions and firm accounts from Chile, we conduct three quantitative exercises that address three distinct questions. The first question asks how much of the quantity discounts observed in the data can be explained by the price-discrimination model presented in Section 4. We do not calibrate parameters to match observed quantity discounts directly. Instead, we calibrate the model to moments of the firm-size distribution and then ask, given this calibration, how much of the observed quantity discounts the model can rationalize.

The second question examines the aggregate welfare implications of price discrimination: what is the effect of allowing firms to set nonlinear prices, and how would welfare change under a policy banning all forms of price discrimination? To answer this, we calibrate the model developed in Section 4 under nonlinear pricing, then we perform a counterfactual experiment in which we impose a ban on price discrimination while holding model parameters fixed and solving for the new equilibrium.

The third question concerns the aggregate welfare implications of market power, which we approach as a measurement exercise. We estimate technology and demand primitives and calibrate the model under two pricing regimes—nonlinear and uniform. Throughout, we refer to these two model specifications as lenses, to emphasize that each represents a distinct way of interpreting

the same economic environment. In this sense, the model serves as an interpretive lens through which we measure and compare the welfare consequences of observed market outcomes.

Quantitatively, the model provides a close fit to the quantity discounts observed in the data. Taking the nonlinear regime as the empirically relevant baseline, a ban on price discrimination would reduce welfare from 0.75 to 0.49 of the efficient benchmark. From a measurement perspective, welfare under nonlinear pricing attains 0.75 of the efficient benchmark, compared with 0.57 under uniform pricing.

5.1 Parameter Estimation

We use Chilean administrative microdata (2005–2022), firm accounts (revenues, wage bill, headcounts, profits, capital), and the universe of firm-to-firm transactions (quantities, prices, counterparties, locations). Parameters tied to technology are measured at fine granularity (6-digit sector by firm type) and mapped to the model's 11 sectors as needed.

Table 3 summarizes the parameters, methods, and granularities used.

Table 3: Estimated Parameters

Parameter	Strategy	Granularity
Labor output elasticity (α_s)	Calibrated from data	626 sectors × firm type
Final demand elasticity (θ_r)	Calibrated from data	626 sectors
Input-output elasticity (θ_{iu})	Calibrated from data	626 sectors × firm type
Final demand-bundle elasticity (φ)	Pin down by CES results and data	11 sectors
Material-bundle elasticity (σ)	COVID-19 shock for Chile estimation	11 sectors
Exit rate(δ)	Calibrated from data	626 sectors
Entry cost (c_e)	Pin down by free entry and data	626 sectors × firm type
Productivity Pareto tail (κ)	MLE estimation	11 sectors × firm type

We calibrate the model under two lenses that interpret observed unit prices differently. Under the nonlinear-pricing lens, two-part tariffs imply that average unit prices converge to marginal prices as quantities rise; we therefore construct moments on a large-firm subsample, where flat-fee dilution makes observed unit prices close to marginal (allocative) prices. Under the uniform-pricing lens, per-unit prices are treated as marginal and quantity-invariant within seller-product-time cells; moments are computed on the full firm population. Appendix D details each estimator and reports results under both lenses.

Labor output elasticity (α_s). This parameter is the Cobb-Douglas weight on primary inputs (labor plus the user cost of capital) in production. We recover it from firm accounts as the nonmaterial

cost share at the 6-digit sector and firm-type level, restricting to large firms to align observed unit prices with marginal prices and winsorizing extremes for stability. Because flat fees are small for these firms, variable-cost shares are reliable proxies for total-cost shares. α_s governs how sectoral output responds to wages relative to materials prices: higher α_s amplifies labor-market importance on aggregate welfare relative to inputs from other other firms.

Final demand output elasticity (θ_r). For large firms, they can be interpreted as Cobb-Douglas weights that allocate the representative consumer's expenditure across retail sectors. With uniform pricing to final consumers, retailer revenues identify sectoral expenditure; we form each sector's share of aggregate retail sales using large firms and average across years. These shares anchor the final-demand system and the welfare accounting used in counterfactuals.

Input-output elasticity (θ_{iu}). For large firms, they can be interpreted as buyer-facing expenditure shares on upstream seller sectors within the materials bundle. Using transaction-level data, we compute for each buyer the fraction of variable materials spending sourced from each upstream sector, aggregate to 6-digit industries within year, and average over time. The resulting matrix provides the micro foundation of the input-output network, determining exposure patterns and the scope for intensive-margin substitution when relative prices move.

Final demand-bundle elasticity of substitution (φ). This parameter is the elasticity of substitution across retail varieties within a sector and, under uniform pricing, coincides with the inverse markup. We recover it from sectoral accounts implied by CES demand, linking pooled sectoral sums of profits for large retailers and averaging across years. Higher φ indicates keener competition and smaller allocative distortions; lower φ sustains higher markups and larger deadweight losses.

Material-bundle elasticity (σ). This parameter measures how easily buyers substitute across varieties within an upstream seller sector in response to relative marginal price changes. We exploit the quasi-experimental disruptions from Chile's early COVID-19 lockdowns (March 2020) by instrumenting the main preshock supplier's relative price change with that supplier's lockdown exposure, estimating sector-specific elasticities via two-stage least squares on 12-month differences, focusing on large buyers, and excluding cases with potentially confounded exposure (buyer location, buyer customers, or other inputs locked down).[25] A higher value of σ implies greater substitutability across suppliers and therefore stronger intensive-margin reallocation. Conversely, a lower σ indicates weaker substitutability, leading to more-persistent and less-flexible buyer–seller relationships.

[25]For sectors with estimates below one, we conservatively adopt the smallest value above one from other sectors.

Exit rate ()δ). This parameter is the one-year hazard that an active firm ceases operations. We compute it at the 6-digit sector interacted with firm-type level by tracking the share of firms present in a given year that are not observed in the following year, and then averaging over 2005–2022. This object disciplines the expected lifespan of an entrant and, together with the discount rate, determines how quickly future profits are attenuated. Higher δ raises the payoff required to justify entry, thins steady-state firm mass for given fundamentals, and shifts the balance between churn and scale. Sectors with elevated exit rate display a larger role for extensive-margin adjustments.

Entry cost (c_e). Entry costs are the labor-measured sunk resources required to create an operating firm. We combine sector-type averages of accounting profits and wages with the empirically estimated exit rates and a standard discount rate to obtain the expected present value of a surviving firm; free entry equates that value, scaled by the share of positive-profit firms, to the labor cost of entry. We report both currency units and "wage-bill equivalents" for comparability across sectors. Higher c_e depresses equilibrium firm mass and raises average scale, sharpening how nonlinear pricing interacts with the extensive margin and rent allocation across links.

Productivity Pareto tail (κ). This parameter governs the thickness of the upper tail of the firm productivity distribution. The model implies a theoretical relationship between the productivity tail parameter κ, the input tail parameter ξ, and the elasticity of substitution σ (or φ), given by

$$\xi = \frac{\kappa}{\sigma - 1}$$

We estimate the input Pareto tail exponent for firm employment using maximum likelihood on the upper tail of the firm size distribution, measured in terms of labor, then we map it to productivity using the model's monotonic link between productivity and input use. The parameter ξ plays a central role in determining allocative efficiency under nonlinear pricing: it governs the extent to which nonlinear contracts mitigate the allocative distortions that would arise under linear uniform pricing. In particular, allocative markups in this model are given by $\rho = \xi\sigma$, whereas under uniform pricing they depend solely on σ.

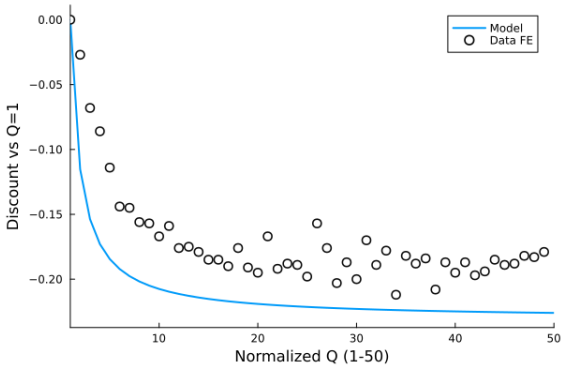
Together, these objects pin down (i) how easily buyers and consumers reallocate across varieties (σ, φ), (ii) how strongly sectors load on wages versus materials costs (α), (iii) who buys what from whom and in what proportion (θ), (iv) how many firms enter and survive (c_e and δ), and (v) how dispersed productivity is within sectors (κ). This configuration determines the balance between intensive reallocation and extensive entry/exit in the quantification and the extent to which nonlinear pricing redistributes surplus without distorting marginal allocations.

5.2 Quantity Discounts: Model vs. Data

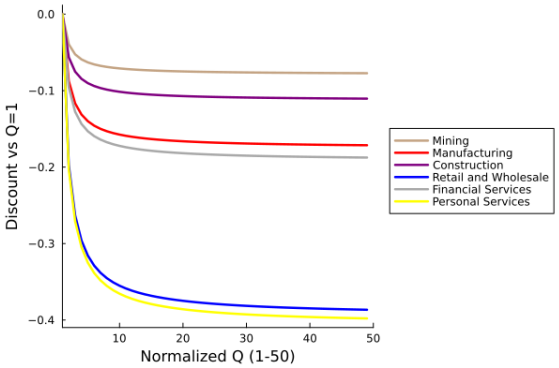
After calibration, we assess empirical validity by confronting the model's quantity-dependent unit prices with the data. Figure 6 contrasts the model-implied quantity discounts with our fixed-effects estimates from the data section. Panel A plots unit prices charged by the average upstream firm to retailers as a function of purchased quantity. We construct the quantity axis by evaluating the model's optimal schedule $q(z)$ of the z distribution and rescaling $q(z)$ monotonically to the interval $[1, 50]$. The "average upstream firm" aggregates seller-product schedules using sector final-demand weights and sales weights over seller-product-time cells. Panel B fixes the seller to be in the Retail and Wholesale sector (the sector with the largest transaction counts) and plots the model implied unit-price schedules to selected buyer sectors under the same quantity normalization as in Panel A.

Figure 6: Calibrated Model Unit Prices

Panel A. Calibrated Model Unit Prices vs. Data Fixed-Effects Regression



Panel B. Model Retail and Wholesale Sector Unit Prices to Selected Buyer Sectors



Notes: Panel A contrasts model-implied quantity discounts with data-based estimates from a within-cell fixed-effects regression (controlling for seller, buyer-group, product, and time). The model curve reports the *average upstream firm*'s unit-price schedule to retailers; we rescale to 50 bins so that $q \in [1, 50]$. The y-axis shows the percentage deviation of the unit price relative to q=1. Panel B fixes the seller to the Retail and Wholesale sector and plots model unit-price schedules to selected buyer sectors under the same normalization.

Quantity-discount moments were not targeted in calibration; parameters were pinned down by independent technology and market-share moments. Nevertheless, the model reproduces the negative unit price–quantity gradient observed in the data. Across buyer sectors within the Retail and Wholesale seller sector, the model captures level shifts in the schedule, consistent with the heterogeneous nonlinear pricing across buyer sectors that we observed in the data.

5.3 Counterfactual: Ban on Price Discrimination

Motivated by pervasive quantity-dependent and group-specific pricing in the data, we study a regulatory counterfactual that bans price discrimination: upstream firms must charge a single, quantity-invariant per-unit price to all buyers (no fixed fees, no buyer- or sector-specific prices, no quantity schedules). We implement the ban by holding the primitives at the nonlinear price calibration—technology, demand, and firm-type distributions fixed—and re-solving the general equilibrium with the contract set restricted to uniform linear pricing. Retail-to-consumer margins are held fixed to isolate the firm-to-firm transaction channel. All endogenous objects (upstream markups to retailers and to upstream buyers, firm entry and masses, and price indices) adjust to their new equilibrium values, yielding a comparison between the nonlinear-pricing benchmark, the uniform-pricing counterfactual, and the efficient (marginal-cost) allocation. Selection is inactive (full participation): with Pareto-distributed productivities and no operating fixed costs, every type participates under both regimes; only equilibrium masses adjust.[26]

We organize the counterfactual analysis in three steps. First, we report the aggregate welfare effect of a ban on price discrimination relative to the nonlinear-pricing baseline; that is, $W^{\text{Uni}}/W^{\text{NLP}}$. Second, we decompose this ratio into intensive and extensive components using the accounting identity in Equation (16) that follows from Proposition 2. The intensive margin captures regime differences in upstream markups to retailers and to other upstream buyers, $\left(\mu_s^{ur}, \mu_s^{uu}\right)$, while the extensive margin captures equilibrium changes in the masses of retail and upstream firms, $\left(N_s^r, N_s^u\right)$, both weighted by final-demand exposure weights. Third, we further decompose each margin by sector and firm type (upstream vs. retail) to identify where welfare gains and losses are concentrated.

In both regimes, labor allocated to entry must increase relative to the efficient benchmark, since positive markups depress labor demand and thereby release labor for entry. By the same logic, labor used for entry must decline under nonlinear pricing relative to the uniform-pricing benchmark. Nonetheless, the overall extensive-margin effect on welfare is a priori ambiguous, as final-demand exposure varies across sectors.

$$\frac{W^{\text{NLP}}}{W^{\text{Uni}}} = \underbrace{\prod_{s\in S}\left(\frac{\mu_s^{ur,\text{NLP}}}{\mu_s^{ur,\text{Uni}}}\right)^{-\tilde{\lambda}_s^{ru}} \prod_{s\in S}\left(\frac{\mu_s^{uu,\text{NLP}}}{\mu_s^{uu,\text{Uni}}}\right)^{-\tilde{\lambda}_s^{uu}}}_{\text{Intensive Margin}} \times \underbrace{\prod_{s\in S}\left(\frac{N_s^{r,\text{NLP}}}{N_s^{r,\text{Uni}}}\right)^{\frac{\theta_s}{\varphi_s-1}} \prod_{s\in S}\left(\frac{N_s^{u,\text{NLP}}}{N_s^{u,\text{Uni}}}\right)^{\frac{\tilde{\lambda}_s^{uu}}{\sigma_s-1}}}_{\text{Extensive Margin}} \tag{16}$$

Aggregate welfare. If a ban on price discrimination forced upstream firms to charge a single, quantity-invariant per-unit price, aggregate welfare would fall from 0.748 in the baseline (non-

[26] A natural variant bans second-degree (nonlinear) pricing while permitting third-degree (buyer-group–specific) linear prices. We do not pursue it here because (i) antitrust practice typically targets both forms and (ii) both are needed to rationalize the observed schedules.

linear pricing) to 0.486 under the ban (Table 4), both relative to the efficient benchmark. This corresponds to a 35% decline in welfare $\left(1 - 0.486/0.748\right)$. Equivalently, the efficiency shortfall would widen from 25.2 to 51.4 percentage points, roughly doubling the distance from efficiency.

Table 4: Aggregate Welfare, Decomposition, and Firm Masses (relative to efficiency)

Price Regime	$\mathcal{W}/\mathcal{W}^{\text{Eff}}$	Intensive	Extensive	Upstream Mass	Retail Mass	$\mathcal{W}^{\text{Uni}}/\mathcal{W}^{\text{NLP}}$
Nonlinear (NLP)	0.748	0.67 (79%)	1.12 (21%)	1.18	1.17	0.650
Uniform (Uni)	0.486	0.46 (93%)	1.06 (7%)	1.00	1.44	

Notes: NLP uses two-part tariffs with buyer-specific pricing, Uni imposes a single per-unit price (no fixed fees; no buyer-specific prices), and Eff implements marginal-cost pricing. Entry and firm masses are re-solved in each regime. *Intensive* captures markup accumulation along supply chains; *Extensive* captures entry (variety) effects via free entry. By construction, Intensive × Extensive = $\mathcal{W}/\mathcal{W}^{\text{Eff}}$, and the shares in parentheses are fractions of absolute log contributions: $|\log(\text{Intensive})|/\left(|\log(\text{Intensive})| + |\log(\text{Extensive})|\right)$ and analogously for the extensive component. Mass ratios are $N^{u,R}/N^{u,\text{Eff}}$ and $N^{r,R}/N^{r,\text{Eff}}$.

Aggregate intensive vs. extensive contributions and entry responses. We quantify how the ban operates relative to efficiency through two margins: an intensive margin that summarizes allocative markups from upstream markups propagated by final-demand exposure along the supply chain, and an extensive margin that summarizes general-equilibrium entry responses and the implied masses of active firms. The factors and shares reported in Table 4 are computed from Proposition 2 using the accounting identity in Equation (16); shares (in parentheses) are fractions of absolute log contributions, since intensive and extensive forces can move in opposite directions.[27]

Two patterns stand out. First, intensive distortions dominate welfare losses relative to efficiency in both pricing regimes: under nonlinear pricing, 79% of the absolute deviation is intensive, and under uniform pricing it rises to 93%. Second, the extensive margin is welfare-improving in both counterfactual but modest: factors exceed one in both regimes (Extensive = 1.12 under NLP and 1.06 under Uni), yet they are quantitatively too small to offset the larger intensive losses, especially under the ban.

Entry responses mirror these patterns. Under nonlinear pricing, firm masses rise relative to efficiency on firm types (Upstream = 1.18, Retail = 1.17), reflecting higher equilibrium expected profits when part of the rent extraction is collected via fixed fees that do not distort marginal conditions. Under uniform pricing, upstream mass is essentially unchanged (Upstream = 1.00), while retail mass expands sharply (Retail = 1.44), but this variety effect is insufficient to counteract the stronger allocative distortion created by uniform per-unit markups.

[27] We proportionally scale the reported factors so that their product exactly matches the welfare ratio relative to efficiency in each regime; the level residual is below 0.03.

Nonlinear vs. uniform pricing: opening welfare ratios by sector. Table 5 reports multiplicative sector-margin contributions to the aggregate ratio $W^{\text{NLP}}/W^{\text{Uni}}$ implied by Equation (16). Entries above one indicate that nonlinear pricing yields higher welfare relative to the ban to revert to uniform prices through that sector-margin; entries below one indicate the opposite. The first two columns ("Intensive") capture allocative effects from upstream markups to retail and upstream to upstream links; the next two ("Extensive") capture variety effects from changes in retail and upstream firm masses. The last column is each sector's net contribution, and the "Product over sectors" row approximates the aggregate ratio (the tiny residual vs. Table 4 reflects rounding and exposure normalization). Final-demand exposure weights $\tilde{\lambda}_s^{ru}$ and $\tilde{\lambda}_s^{uu}$ (stats in Appendix E.1) pin down how strongly sector-s markups load into final-demand prices and thus how powerful intensive relief will be.[28]

Table 5: Aggregate Welfare Decomposition by Sector: Nonlinear Relative to Uniform Pricing

Sector	Intensive (allocative)		Extensive (variety)		Net NLP/Uni
	Retailers	Upstream	Retailers	Upstream	
Agriculture	1.010	1.010	0.997	1.005	1.022
Mining	1.003	1.003	0.999	1.014	1.019
Manufacturing	1.024	1.029	0.991	1.002	1.047
Utilities	1.016	1.006	0.996	1.033	1.051
Construction	1.061	1.022	0.980	1.119	1.189
Retail and Wholesale	1.037	1.070	0.992	1.005	1.106
Transport and ICTs	1.007	1.023	0.981	1.000	1.011
Financial Services	1.012	1.008	0.943	0.998	0.960
Real Estate Services	1.009	1.004	0.996	1.023	1.033
Business Services	1.005	1.006	0.989	0.999	0.999
Personal Services	1.001	1.001	0.998	1.000	1.000
Product over sectors	1.197	1.198	0.870	1.207	1.507

Notes: Entries are multiplicative sector-margin contributions to the aggregate ratio $W^{\text{NLP}}/W^{\text{Uni}}$ per Equation (16). *Intensive (Retailers)* and *Intensive (Upstream)* correspond to exposure-weighted allocative components on upstream→retail (I_s^{ur}) and upstream→upstream (I_s^{uu}) links, respectively; *Extensive (Retailers)* and *Extensive (Upstream)* are variety components from retail and upstream masses (E_s^r, E_s^u). Values > 1 indicate that nonlinear pricing raises welfare relative to uniform pricing through that sector-margin; values < 1 indicate the opposite. The sectoral *Net* NLP/Uni equals the product $I_s^{ur} \cdot I_s^{uu} \cdot E_s^r \cdot E_s^u$. The *Product over sectors* row reports $\prod_s I_s^{ur}$, $\prod_s I_s^{uu}$, $\prod_s E_s^r$, and $\prod_s E_s^u$; their product approximates the aggregate ratio in Table 4, with small differences due to rounding and exposure normalization.

Two patterns are immediate on the intensive side. First, higher welfare under nonlinear prices relative to the ban is broad-based: intensive factors exceed one in every sector; nonlinear prices unambiguously improve welfare across all sectors relative to uniform prices on the intensive mar-

[28] In our data, these exposures are concentrated in Retail and Wholesale, Manufacturing, Transport and ICTs, and Construction; consequently, attenuating double marginalization in these sectors delivers outsized intensive gains, while extensive responses are smaller and mixed across sectors.

gin due to attenuated double marginalization. Second, improvements are largest in sectors that are highly exposed to final demand: Construction (Retailers = 1.061, Upstream = 1.022), Retail and Wholesale (1.037, 1.070), Manufacturing (1.024, 1.029), and Utilities (1.016, 1.006). These sectors account for the bulk of the markup-relief advantage of nonlinear pricing.

On the extensive side, the pattern is mixed but small in magnitude. Upstream variety generally expands under nonlinear pricing (e.g., Construction = 1.119, Utilities = 1.033, Real Estate = 1.023, Mining = 1.014), while retail variety often contracts (e.g., Construction = 0.980, Manufacturing = 0.991, Retail and Wholesale = 0.992, Transport and ICTs = 0.981, Financial Services = 0.943). This is consistent with fixed-fee rents shifting expected profits upstream: entry tilts toward upstream sectors while retail entry is less favored. Quantitatively, these variety terms are too small to overturn the intensive gains.

Netting all four columns, nonlinear pricing yields higher welfare relative to uniform pricing in most sectors, with the largest contributions from Construction (1.189) and Retail and Wholesale (1.106), followed by Utilities (1.051) and Manufacturing (1.047). Financial Services is the lone sizable exception (0.960), and Business Services is essentially neutral (0.999). These sectoral patterns mirror the exposure weights emphasized in the decomposition: where final-demand exposure is high, attenuating double marginalization under nonlinear pricing delivers the largest welfare gains relative to uniform pricing.

Policy implications. An across-the-board prohibition of quantity discounts is not warranted unless accompanied by an output subsidy. The welfare gains from nonlinear pricing arise primarily through reductions in intensive allocative distortions in upstream sectors with high final-demand exposure; extensive (entry-variety) effects are positive but quantitatively secondary. Regulation should first evaluate conduct using marginal (allocative) prices rather than average unit prices. If nonlinear pricing is pervasive, regulators should target markup accumulation along highly exposed supply-chain links and constrain rent extraction while preserving low marginal prices and maintaining entry incentives.

5.4 Measurement: Aggregate Welfare, Nonlinear vs. Uniform Lens

We measure aggregate welfare under two pricing lenses, nonlinear and uniform. For the same data, we use lens-specific pricing parameters conditional on the interpretation of observed unit prices. Relative to the efficient (marginal-cost) benchmark, welfare equals 0.748 under the nonlinear lens and 0.565 under the uniform lens (Table 6). The welfare shortfall is 25.2% under nonlinear pricing versus 43.5% under uniform pricing; adopting the nonlinear lens closes 18.3 percentage points, which is about 42% ($0.183/0.435$) of the uniform-lens efficiency gap. This suggests that the aggregate costs of market power are lower if measured with a model that allows for price discrimination.

Table 6: Welfare: Nonlinear vs. Uniform Lenses (relative to efficiency)

Price Lens	$\mathcal{W}^L/\mathcal{W}^{\text{Eff}}$	Intensive	Extensive
Nonlinear	0.748	0.68 (81%)	1.10 (19%)
Uniform	0.565	0.55 (97%)	1.02 (3%)

Notes: *Intensive* captures markup accumulation along supply chains; *Extensive* captures entry (variety) effects via free entry. By construction, Intensive × Extensive = $\mathcal{W}^L/\mathcal{W}^{\text{Eff}}$, and the shares are computed as $|\log(\text{Intensive})|/(|\log(\text{Intensive})| + |\log(\text{Extensive})|)$ and $|\log(\text{Extensive})|/(|\log(\text{Intensive})| + |\log(\text{Extensive})|)$, respectively.

Decomposing the welfare ratios, the nonlinear lens implies an intensive factor of 0.68 and an extensive factor of 1.10, so roughly 81% of the log distance to efficiency comes from the intensive margin and 19% from the extensive margin. Under the uniform lens, the intensive factor is 0.55 and the extensive factor is 1.02, with about 97% of the log distance accounted for by the intensive margin. These accounting facts map directly into the mechanisms that differ across lenses, clarifying why the nonlinear lens delivers higher welfare.

Relative to uniform pricing, under nonlinear pricing via two-part tariffs, surplus extraction shifts toward a combination of a flat fee and per-unit markups: the marginal price moves closer to marginal cost (smaller intensive loss). The induced change in expected profits (entering the free-entry condition) is sign-ambiguous because flat fees reallocate surplus across firms and buyer relationships; in our calibration, average equilibrium expected profits are higher, raising the mass of entrants (Extensive > 1), expanding variety, and lowering sectoral CES price indices. By contrast, under uniform pricing the full rent extraction resides in per-unit markups, distorting all marginal trades, raising price indices, and depressing variable profits; entry thus provides little offset, consistent with the near-unit extensive factor in Table 6.

Taken together, measuring the aggregate welfare cost of market power under the nonlinear lens attenuates allocative distortions at the margin and (in this calibration) strengthens the variety offset, matching the $0.68 \times 1.10 \approx 0.748$ vs. $0.55 \times 1.02 \approx 0.565$ welfare levels under identical technology. Motivated by this quantitative evidence, and given pervasive quantity-dependent and group-specific pricing in the data, we adopt the nonlinear lens as the empirically grounded baseline for policy counterfactuals.

6 Conclusion

Price discrimination in supply chains is central to measuring the aggregate welfare costs of market power and informing current policy debates. Using population-level firm-to-firm transactions from Chile, we document systematic departures from uniform pricing: unit prices decline with quantity purchased and vary across buyer sectors, consistent with both second-degree (quan-

tity discounts) and third-degree (buyer-specific schedules) price discrimination. While we interpret this evidence as indicative, these pervasive patterns motivate our theoretical and quantitative analysis. We develop a multi-sector general equilibrium model where firms simultaneously charge and pay nonlinear prices.

Under standard assumptions—CES demand, Pareto-distributed productivity—optimal contracts take the form of buyer-sector-specific two-part tariffs, a flat fee and a marginal price. Relative to uniform pricing, this structure brings marginal prices closer to marginal cost, mitigating double marginalization along supply chains, while flat fees redistribute profits and affect entry. Calibrating the model to Chilean data, we find that banning price discrimination reduces welfare from 75% to 49% of the efficient benchmark, with losses arising primarily from worsened allocative efficiency in sectors with high final-demand exposure. Entry effects are second-order. When interpreting the same data as uniform rather than nonlinear pricing, measured welfare is 57% versus 75% of the efficient benchmark—standard approaches that assume uniform pricing substantially overstate the welfare costs of market power.

These results carry two key implications. First, broad prohibitions on price discrimination without output subsidies can reduce welfare by forcing firms to distort quantities rather than redistribute rents. Policy should target markup accumulation in supply chains directly rather than restricting pricing instruments. Second, incorporating nonlinear pricing into models of market power is essential for accurate welfare measurement. Our framework provides a practical methodology, separating allocative from entry channels and mapping sector-level distortions to aggregate outcomes through input-output linkages.

References

Acemoglu, D., Carvalho, V. M., Ozdaglar, A., and Tahbaz-Salehi, A. (2012). The network origins of aggregate fluctuations. *Econometrica*, 80(5):1977–2016.

Armstrong, M. (1996). Multiproduct nonlinear pricing. *Econometrica: Journal of the Econometric Society*, pages 51–75.

Autor, D., Dorn, D., Katz, L. F., Patterson, C., and Van Reenen, J. (2020). The fall of the labor share and the rise of superstar firms. *The Quarterly journal of economics*, 135(2):645–709.

Baqaee, D. and Farhi, E. (2020a). Entry vs. rents: Aggregation with economies of scale. Technical report, National Bureau of Economic Research.

Baqaee, D. R. and Farhi, E. (2020b). Productivity and misallocation in general equilibrium. *The quarterly journal of economics*, 135(1):105–163.

Barkai, S. (2020). Declining labor and capital shares. *The Journal of Finance*, 75(5):2421–2463.

Bigio, S. and La'o, J. (2020). Distortions in production networks. *The Quarterly Journal of Economics*, 135(4):2187–2253.

Boehm, J. and Oberfield, E. (2020). Misallocation in the market for inputs: Enforcement and the organization of production. *The Quarterly Journal of Economics*, 135(4):2007–2058.

Bornstein, G. and Peter, A. (2024). Nonlinear pricing and misallocation. Technical report, National Bureau of Economic Research.

Burstein, A., Cravino, J., and Rojas, M. (2024). Input price dispersion across buyers and misallocation. Technical report, Central Bank of Chile.

Carvalho, V. M. and Tahbaz-Salehi, A. (2019). Production networks: A primer. *Annual Review of Economics*, 11(1):635–663.

De Loecker, J., Eeckhout, J., and Unger, G. (2020). The rise of market power and the macroeconomic implications. *The Quarterly journal of economics*, 135(2):561–644.

Dhyne, E., Kikkawa, A. K., Kong, X., Mogstad, M., and Tintelnot, F. (2023). Endogenous production networks with fixed costs. *Journal of International Economics*, 145:103841.

Dhyne, E., Kikkawa, A. K., and Magerman, G. (2022). Imperfect competition in firm-to-firm trade. *Journal of the European Economic Association*, 20(5):1933–1970.

Dupuit, J. (1844). On the measurement of the utility of public works. *International Economic Papers*, 2(1952):83–110.

Edmond, C., Midrigan, V., and Xu, D. Y. (2023). How costly are markups? *Journal of Political Economy*, 131(7):1619–1675.

Hall, R. E. (2018). New evidence on the markup of prices over marginal costs and the role of mega-firms in the us economy. Technical report, National Bureau of Economic Research.

Harberger, A. C. (1954). Monopoly and resource allocation. *The American Economic Review*, 44(2):77–87. Papers and Proceedings of the Sixty-sixth Annual Meeting of the American Economic Association.

Hopenhayn, H. A. (1992). Entry, exit, and firm dynamics in long run equilibrium. *Econometrica: Journal of the Econometric Society*, pages 1127–1150.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and manufacturing tfp in china and india. *The Quarterly journal of economics*, 124(4):1403–1448.

Jones, C. I. (2011). Intermediate goods and weak links in the theory of economic development. *American Economic Journal: Macroeconomics*, 3(2):1–28.

Laffont, J.-J. and Tirole, J. (1993). *A theory of incentives in procurement and regulation*. MIT press.

Maskin, E. and Riley, J. (1984). Monopoly with incomplete information. *The RAND Journal of Economics*, 15(2):171–196.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *econometrica*, 71(6):1695–1725.

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *The review of economic studies*, 38(2):175–208.

Mussa, M. and Rosen, S. (1978). Monopoly and product quality. *Journal of Economic theory*,

18(2):301–317.

Oberfield, E. (2018). A theory of input–output architecture. *Econometrica*, 86(2):559–589.

Quesnay, F. (1894). *Tableau oeconomique*. Macmillan.

Restuccia, D. and Rogerson, R. (2008). Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic dynamics*, 11(4):707–720.

Spence, M. (1977). Nonlinear prices and welfare. *Journal of public economics*, 8(1):1–18.

Stole, L. A. (2007). Price discrimination and competition. *Handbook of industrial organization*, 3:2221–2299.

Tirole, J. (1988). *The theory of industrial organization*. MIT press.

Varian, H. R. (1989). Price discrimination. *Handbook of industrial organization*, 1:597–654.

Wilson, R. B. (1993). *Nonlinear pricing*. Oxford University Press, USA.

Appendix

A Optimal Nonlinear Price Derivation

Primitives. Consider a screening problem in which a monopolist offers quantity–transfer bundles to buyers with private productivity types z, drawn from distribution $F(z)$ with density $f(z)$ and support $[\underline{z}, \infty)$. The seller faces constant marginal cost $c > 0$. The buyer's revenue function[29] is $R(z, q)$, increasing in both arguments and differentiable in z. A contract specifies $(q(z), T(z))$, so type z earns net surplus

$$\Pi(z) = R(z, q(z)) - T(z).$$

Seller problem. The seller chooses a menu $\{q(z), T(z)\}$ to maximize expected profit

$$\max_{\{q(z),T(z)\}} \int_{\underline{z}}^{\infty} \Big[T(z) - cq(z)\Big] f(z)\, dz,$$

subject to individual rationality (IR) and incentive compatibility (IC):

$$\Pi(z) \geq 0, \qquad \Pi(z) \geq R(z, q(\tilde{z})) - T(\tilde{z}) \quad \forall z, \tilde{z} \geq \underline{z}.$$

We assume monotone allocations $q'(z) \geq 0$, so higher types purchase weakly more.

Envelope and transfers. By the Envelope Theorem,

$$\Pi'(z) = \frac{\partial R(z, q(z))}{\partial z}, \qquad \Pi(\underline{z}) = 0,$$

so

$$\Pi(z) = \int_{\underline{z}}^{z} \frac{\partial R(s, q(s))}{\partial s}\, ds, \quad T(z) = R(z, q(z)) - \Pi(z).$$

Virtual surplus. Substituting $T(z)$ into the seller's objective and exchanging the order of integration yields

$$\Pi_{\text{seller}} = \int_{\underline{z}}^{\infty} \Big[R(z, q(z)) - \tfrac{1-F(z)}{f(z)} \tfrac{\partial R(z,q(z))}{\partial z} - cq(z)\Big] f(z)\, dz.$$

Define the virtual surplus

$$\phi(z, q) = R(z, q) - \frac{1}{h(z)} \frac{\partial R(z, q)}{\partial z}, \qquad h(z) = \frac{f(z)}{1 - F(z)},$$

so that

$$\Pi_{\text{seller}} = \int_{\underline{z}}^{\infty} \Big[\phi(z, q(z)) - cq(z)\Big] f(z)\, dz.$$

[29] If buyers are final consumers, $R(z, q)$ can be interpreted as gross utility.

Virtual surplus adjusts revenues for the information rents needed to preserve truthful revelation; the inverse hazard rate $1/h(z)$ scales those rents.

Functional forms. We now impose the functional forms used in the main text. Let types be Pareto with shape $\kappa > 1$ and, without loss, lower bound $\underline{z} = 1$:

$$f(z) = \kappa z^{-\kappa-1}, \qquad F(z) = 1 - z^{-\kappa}, \qquad h(z) = \frac{\kappa}{z}.$$

Let revenues be homogeneous and normalized:

$$R(z,q) = z^{\frac{\sigma-1}{\sigma}} q^{\frac{\sigma-1}{\sigma}}, \qquad \sigma > 1.$$

Then

$$\phi(z,q) = z^{\frac{\sigma-1}{\sigma}} q^{\frac{\sigma-1}{\sigma}} \left(1 - \frac{\sigma-1}{\kappa\sigma}\right) = z^{\frac{\sigma-1}{\sigma}} q^{\frac{\sigma-1}{\sigma}} \left(\frac{\rho-1}{\rho}\right), \qquad \rho \equiv \frac{\sigma\kappa}{\sigma-1}.$$

Allocation. The integrand is pointwise concave in q, so the optimal $q(z)$ solves

$$\max_{q(z)} \left\{\frac{\rho-1}{\rho} z^{\frac{\sigma-1}{\sigma}} q^{\frac{\sigma-1}{\sigma}} - cq\right\}.$$

The FOC yields

$$\frac{\rho-1}{\rho} \cdot \frac{\sigma-1}{\sigma} z^{\frac{\sigma-1}{\sigma}} q(z)^{-\frac{1}{\sigma}} = c \quad \Rightarrow \quad q(z) = K z^{\sigma-1}, \qquad K = \left[\frac{1}{c}\frac{\rho-1}{\rho}\frac{\sigma-1}{\sigma}\right]^{\sigma}.$$

Because $\kappa > 1$, then $\rho > 1$, the coefficient $(\rho-1)/\rho > 0$ and the solution is interior for all z. Moreover, if $\kappa > \sigma - 1$ the integrated objective is finite since the profit integrand scales like $z^{\sigma-\kappa-2}$ under the Pareto tail.

Two-part tariff implementation. Consider an indirect mechanism with tariff $T(q)$. A type z chooses q to satisfy

$$T'(q) = \frac{\partial R}{\partial q}(z,q) = \frac{\sigma-1}{\sigma} z^{\frac{\sigma-1}{\sigma}} q^{-\frac{1}{\sigma}}.$$

Evaluated at the target allocation $q(z) = Kz^{\sigma-1}$,

$$\frac{\partial R}{\partial q}\Big(z, q(z)\Big) = \frac{\sigma-1}{\sigma} z^{\frac{\sigma-1}{\sigma}} \Big(Kz^{\sigma-1}\Big)^{-\frac{1}{\sigma}} = \frac{\sigma-1}{\sigma} K^{-\frac{1}{\sigma}},$$

which is independent of z. Hence $T'(q) \equiv p^{NLP}$ is constant on the implemented range and

$$T(q) = F + p^{NLP} q.$$

To pin down p^{NLP}, note that under a linear marginal price p,

$$q^{\mathrm{BR}}(z;p) = \arg\max_{q} \left\{ z^{\frac{\sigma-1}{\sigma}} q^{\frac{\sigma-1}{\sigma}} - pq \right\} = \left(\frac{\sigma-1}{\sigma p} \right)^{\sigma} z^{\sigma-1}.$$

Equating $q^{\mathrm{BR}}(z;p) = Kz^{\sigma-1}$ gives $p^{NLP} = \frac{\sigma-1}{\sigma} K^{-1/\sigma}$. Using the expression for K,

$$p^{NLP} = \frac{\rho}{\rho-1}\, c.$$

Choose the flat fee to bind the lowest-type IR:

$$F = R\Big(1, q(1)\Big) - p^{NLP} q(1), \qquad q(1) = K.$$

Equivalently, since $q\,\partial R/\partial q = \frac{\sigma-1}{\sigma} R$, one can write $F = \frac{p^{NLP}}{\sigma-1}\, q(1)$.

Result. With Pareto types and $R(z,q) = z^{\frac{\sigma-1}{\sigma}} q^{\frac{\sigma-1}{\sigma}}$, the Mirrlees-optimal allocation $q(z) = Kz^{\sigma-1}$ is implemented by a single two-part tariff

$$T(q) = F + p^{NLP} q, \qquad p^{NLP} = \frac{\rho}{\rho-1} c, \quad \rho = \frac{\sigma\kappa}{\sigma-1},$$

with F chosen to satisfy the bottom-type IR. All types are served (since $\rho > 1$), IC and IR hold, and the indirect mechanism coincides with the solution to the screening problem.

Information-adjusted revenue in closed form. Plugging the optimal $q(z)$ into $\phi(z,q)$ and integrating over types yields

$$\int_1^{\infty} \phi\Big(z, q(z)\Big) f(z)\, dz = \frac{\kappa}{\kappa-\sigma+1} \left(\frac{\sigma\kappa-\sigma+1}{\sigma\kappa} \right)^{\sigma} \left(\frac{\sigma-1}{\sigma} \right)^{\sigma-1} \frac{1}{c^{\sigma-1}},$$

which exists if and only if $\kappa > \sigma - 1$. This closed form shows the information-adjusted revenue is homogeneous of degree $1-\sigma$ in marginal cost and depends on primitives only through (σ, κ, c).

Seller revenue is homogeneous in q. The virtual surplus can also be rewritten as:

$$\phi(z,q) \;=\; R(z,q) - \frac{1}{h(z)} \frac{\partial R(z,q)}{\partial z} \;=\; \Big(1 - \frac{\sigma-1}{\sigma\,\kappa}\Big) z^{\frac{\sigma-1}{\sigma}}\, q^{\frac{\sigma-1}{\sigma}}.$$

Hence, for any $t > 0$, $\phi(z, tq) = t^{\frac{\sigma-1}{\sigma}} \phi(z,q)$: the seller-side revenue term is homogeneous in q with degree $(\sigma-1)/\sigma$.

Aggregating over types (support $[1,\infty)$), the seller's revenue functional is

$$\mathcal{R}[q] \;=\; \int_1^\infty \phi\Big(z, q(z)\Big)\, f(z)\, dz \;=\; \frac{\sigma\kappa-\sigma+1}{\sigma} \int_1^\infty z^{\frac{\sigma-1}{\sigma}-\kappa-1}\, q(z)^{\frac{\sigma-1}{\sigma}}\, dz,$$

so for any $t>0$ we have $\mathcal{R}[tq] = t^{\frac{\sigma-1}{\sigma}}\mathcal{R}[q]$. This gives the full expression for the seller's revenue and makes its homogeneity in q explicit.

Expected revenue by source. Aggregating across buyers, total revenue equals the sum of flat fees and per-unit revenue:

$$\text{Total Revenue} = \underbrace{F\int_1^\infty f(z)\,dz}_{\text{flat fees}} \;+\; \underbrace{\Big(\int_1^\infty q(z)\,f(z)\,dz\Big)\,p^{\text{NLP}}}_{\text{per-unit revenue}}.$$

With $\int_1^\infty f(z)\,dz = 1$ and $\int_1^\infty z^{\sigma-1}f(z)\,dz = \frac{\kappa}{\kappa-\sigma+1}$ (for $\kappa>\sigma-1$), and using

$$q(1) = \left[\frac{1}{c}\cdot\frac{\sigma-1}{\sigma}\cdot\frac{\sigma\kappa-\sigma+1}{\sigma\kappa}\right]^\sigma, \qquad F = q(1)^{\frac{\sigma-1}{\sigma}} - p^{\text{NLP}}\,q(1), \qquad p^{\text{NLP}} = \frac{\sigma\kappa}{\sigma\kappa-\sigma+1}\,c,$$

Let

$$B \;\equiv\; q(1) = \left[\frac{1}{c}\cdot\frac{\sigma-1}{\sigma}\cdot\frac{\sigma\kappa-\sigma+1}{\sigma\kappa}\right]^\sigma, \qquad p^{\text{NLP}} = \frac{\sigma\kappa}{\sigma\kappa-\sigma+1}\,c$$

Total revenue aggregates flat fees and per-unit revenue:

$$\text{Total Revenue} = \underbrace{\Big[B^{\frac{\sigma-1}{\sigma}} - p^{\text{NLP}}\,B\Big]}_{\text{flat fees}} \;+\; \underbrace{p^{\text{NLP}}\,B\,\frac{\kappa}{\kappa-\sigma+1}}_{\text{per-unit revenue}}.$$

The first bracketed term is total flat-fee revenue, while the second term is per-unit markup revenue proportional to the Pareto moment $\mathbb{E}[z^{\sigma-1}] = \kappa/(\kappa-\sigma+1)$; hence heterogeneity ($\kappa$) and technology/costs ($\sigma, c$ through B) shift levels but not the two-part structure.

A.1 Virtual Surplus and Full Participation

The monopolist optimally serves all buyer types, including the lowest type $z=1$. The logic is transparent in virtual-surplus form. With homogeneous revenue,

$$R(z,q) = z^{\frac{\sigma-1}{\sigma}}\,q^{\frac{\sigma-1}{\sigma}}, \qquad \sigma>1,$$

and Pareto types with hazard $h(z) = \kappa/z$, we have

$$\frac{\partial R(z,q)}{\partial z} = \frac{\sigma-1}{\sigma} z^{-\frac{1}{\sigma}} q^{\frac{\sigma-1}{\sigma}}, \qquad \phi(z,q) = z^{\frac{\sigma-1}{\sigma}} q^{\frac{\sigma-1}{\sigma}} \Big(1 - \frac{\sigma-1}{\kappa\sigma}\Big) = z^{\frac{\sigma-1}{\sigma}} q^{\frac{\sigma-1}{\sigma}} \frac{\rho-1}{\rho}.$$

Hence the net contribution of type z at allocation $q(z)$ is

$$\mathrm{VS}(z) = \phi(z, q(z)) - cq(z).$$

Evaluated at the lowest type,

$$\mathrm{VS}(1) = q(1)^{\frac{\sigma-1}{\sigma}} \frac{\rho-1}{\rho} - cq(1),$$

which is strictly positive whenever $\rho > 1$ (equivalently, $\kappa > 1$). Since low types are abundant under Pareto, excluding them lowers profits: the mass of low types more than compensates their low individual surplus. The exclusion trade-off therefore resolves in favor of full participation.

A.2 No Profitable Price Deviation

We show that the seller cannot profit by deviating from the optimal nonlinear schedule and charging a different per-unit price for a given quantity. This validates the two-part tariff with constant unit price $p^{NLP} = \frac{\rho}{\rho-1} c$ as seller-optimal.

A buyer facing marginal price $p(q)$ solves

$$\max_q \{R(z,q) - T(q)\}, \qquad T'(q) = p(q).$$

With $R(z,q) = z^{\frac{\sigma-1}{\sigma}} q^{\frac{\sigma-1}{\sigma}}$,

$$\frac{\partial R}{\partial q} = \frac{\sigma-1}{\sigma} z^{\frac{\sigma-1}{\sigma}} q^{-\frac{1}{\sigma}} = p(q).$$

Solving for the type indifferent at (q,p) yields the inverse demand for the q-th unit:

$$z(q,p) = \Big(\frac{\sigma}{\sigma-1}\, p\Big)^{\frac{\sigma}{\sigma-1}} q^{\frac{1}{\sigma-1}}.$$

If the seller posts an alternative price p for quantity q, only types $z \geq z(q,p)$ purchase that unit, so demand is

$$D(q,p) = 1 - F\Big(z(q,p)\Big) = z(q,p)^{-\kappa} \quad \text{(Pareto)}.$$

Profit from this deviation is

$$\pi(q,p) = D(q,p)\,(p-c) = \left[\Big(\frac{\sigma}{\sigma-1}\, p\Big)^{\frac{\sigma}{\sigma-1}} q^{\frac{1}{\sigma-1}}\right]^{-\kappa} (p-c).$$

Maximizing w.r.t. p yields the first-order condition whose solution is

$$\frac{p}{c} = \frac{\rho}{\rho - 1}, \qquad \rho = \frac{\sigma\kappa}{\sigma - 1},$$

i.e., $p = p^{NLP}$. Any unilateral price deviation reduces profit. Thus the nonlinear price is robust to such deviations and coincides with the allocative price embedded in the mechanism (cf. Wilson (1993)).

Figure A1: No Profitable Price Deviation

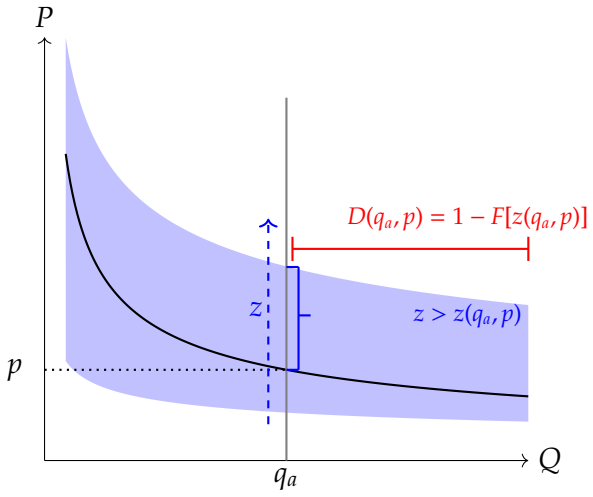


Figure A1 illustrates the logic: raising p at a given q_a shrinks the set of buyers above the new cutoff $z(q_a, p)$; the loss in volume offsets the price gain unless $p = p^{NLP}$.

B Additional Descriptive Evidence

B.1 Residual Price Determinants by Selected Industries

To assess whether the pattern of nonlinear pricing driven by buyer observables generalizes across sectors, we replicate the residual decomposition analysis for the two industries with the highest volume of transactions: Manufacturing and Retail and Wholesale. For both sectors, we estimate the following regression:

$$\ln p_{igjt} = \beta_0 + \Psi_{igmS} + \epsilon_{ijgt}, \tag{17}$$

where p_{igjt} is the unit price of a product g sold by seller i to buyer j at time t, and Ψ_{igmS} represents seller-product-month fixed effects interacted with different sets of quantity and buyer-

side controls S. Buyer groups B are defined based on 11 sectors, 3 firm-size categories, and 16 regions.

Table A1: Price residual determinants: Manufacturing

	(1)	(2)	(3)	(4)
R^2	0.581	0.389	0.312	0.776
S = Quantity	✓			
S = Buyer		✓		
S = Buyer Group			✓	
S = Quantity × Buyer Group				✓
N	136M	136M	136M	136M

Table A2: Price residual determinants: Retail and Wholesale

	(1)	(2)	(3)	(4)
R^2	0.296	0.391	0.309	0.471
S = Quantity	✓			
S = Buyer		✓		
S = Buyer Group			✓	
S = Quantity × Buyer Group				✓
N	180M	180M	180M	180M

In both sectors, the pattern remains unchanged: quantity discounts (second-degree price discrimination) and buyer group-based pricing (third-degree) explain the majority of price dispersion once product and time effects are controlled for. This reinforces our main finding that nonlinear prices shaped by buyer-side observables are a pervasive feature of pricing in supply chains.

B.2 Average Quantity Discount by Sector

Table A3: Average Quantity Discount by Sector

Sector	Mean Q discount	N transactions
All sectors	-0.042	430M
Agriculture	-0.042	2M
Mining	-0.016	1M
Manufacturing	-0.036	118M
Utilities	0.000	6M
Construction	-0.129	1M
Retail and Wholesale	-0.048	270M
Transport & ICTs	-0.032	12M
Financial Services	-0.002	49M
Real Estate Services	-0.052	1M
Business Services	-0.089	5M
Personal Services	-0.053	1M

B.3 Test for Buyer Power Data Generation Process

To examine whether observed quantity discounts reflect buyer power rather than seller-driven price discrimination, we exploit cross-sectional variation in the number of suppliers each buyer transacts with during the sample period. The underlying idea is that buyers with access to a larger number of sellers may possess stronger outside options, enhancing their bargaining position and enabling them to negotiate better pricing terms. We define buyer power as the logarithm of the total number of distinct sellers each buyer purchases from within the observed month. We then test whether buyer power flattens quantity discounts by estimating the interaction between log quantity and buyer power in a log-linear price regression. Specifically, we estimate:[30]

$$\ln p_{igjt} = \beta_0 + \beta_1 \ln q_{igjt} + \beta_2 \left(\log q_{ijt} \times \log \text{NumProviders}_j \right) + \Psi_{igm} + \epsilon_{ijgt},$$

A positive coefficient on the interaction term ($\beta_2 > 0$) would suggest that quantity discounts become flatter as buyer power increases, consistent with buyers using their broader supplier base to resist steep discounts or nonlinear price schedules.

[30] Standard errors are clustered at the buyer level to account for within-buyer correlation.

We find that $\beta_1 = -0.0462$ and $\beta_2 = -0.0098$, both estimated with standard errors below 0.0001. While the interaction term is statistically significant, the magnitude is economically negligible. This suggests that buyer power, as measured by the number of suppliers, does not appear to be the primary mechanism generating quantity discounts. If anything, the evidence is more consistent with seller-driven price discrimination rather than buyer power shaping quantity discounts.

B.4 Firm Sales Partition

We find that firms in Chile have a clear partition on firms' buyers: 79% of firms weighted by sales sell all their output either to only other firms (67%) or to only final consumers (12%). As we can combine firm-to-firm transaction data with firms' accounting variables, we build an indicator variable that takes the value of 0 if all firm sales go to final consumers and 1 if sales go only to other firms, and we weigh the indicator by firm sales.

Table A4: Firms sales partition

Sector (Supply Chain Transactions Value Share)	**All to final consumer**	**All to other firms**
Firm population (100%)	0.12	0.67
Agriculture (2%)	0.05	0.60
Mining (1%)	0.27	0.08
Manufacturing (15%)	0.06	0.69
Utilities (3%)	0.20	0.52
Construction (8%)	0.02	0.89
Retail and Wholesale (32%)	0.09	0.69
Transport and ICTs (10%)	0.16	0.68
Financial Services (18%)	0.18	0.68
Real Estate Services (1%)	0.25	0.38
Business Services (7%)	0.09	0.81
Personal Services (2%)	0.69	0.10

Notes: Exports are excluded. The remaining 16% of sales shares for the firm population are firms that sell to both final consumers and other firms. We observe firm-to-firm sales an fims total sales, we compute the sales to consumer as the residual between both. For 2% of firms, we get negative sales to consumers and exclude them from this table.

As shown in Table A4, there is heterogeneity across sectors, though the partition between firms

selling to final consumers and other firms is present across all sectors.

C Model Details and Derivations

C.1 Verification for Retailers

Within retail sector s, CES demand over differentiated varieties implies

$$y_j \;=\; Y_s\Big(\frac{p_j}{P_s}\Big)^{-\varphi_s}, \qquad Y_s = \theta_s Y, \quad P_Y \equiv 1,$$

hence the inverse demand $p_j = P_s\,(y_j/Y_s)^{-1/\varphi_s}$. Revenue as a function of own output Q_j is

$$R_j \;=\; p_j Q_j \;=\; P_s\,(\theta_s Y)^{1/\varphi_s}\,Q_j^{(\varphi_s-1)/\varphi_s}.$$

Therefore, for retailers ($\ell = r$) the revenue guess for retailers holds with

$$\psi_s^r = \frac{\varphi_s - 1}{\varphi_s}, \qquad A_s^r = P_s\,(\theta_s Y)^{1/\varphi_s}.$$

C.2 Verification for Upstream Sellers (Homothetic Revenue)

Fix a seller sector s' with elasticity $\sigma_{s'} > 1$ and a seller $j \in \mathcal{U}_{s'}$ with marginal cost $c_j > 0$. For buyers in partition (ℓ, s), the two–part tariff of Proposition 1 implies the allocative price $p_{js}^{\ell} = \mu_{ss'}^{\ell} c_j$ with $\mu_{ss'}^{\ell} = \rho_{ss'}^{\ell}/(\rho_{ss'}^{\ell} - 1)$ and $\rho_{ss'}^{\ell} = \xi_s^{\ell}\sigma_{s'} > \sigma_{s'}$.

Step 1: quantity aggregation. By the CES share rule, for buyer $i = (\ell, s, z_i)$,

$$m_{ij} \;=\; M_{is'}\left(\frac{p_{js}^{\ell}}{P_{ss'}^{\ell}}\right)^{-\sigma_{s'}}.$$

Let $\tilde{\nu}_{\ell s}$ be the buyer distribution in (ℓ, s) normalized to one and define the average sector–s' bundle per buyer

$$\widehat{D}_{ss'}^{\ell} \;\equiv\; \int M_{is'}\, d\tilde{\nu}_{\ell s}(i).$$

Total quantity sold by j to partition (ℓ, s) is then

$$Q_j^{\ell,s} \;=\; N_s^{\ell}\,\widehat{D}_{ss'}^{\ell}\left(\frac{p_{js}^{\ell}}{P_{ss'}^{\ell}}\right)^{-\sigma_{s'}}.$$

Summing over (ℓ, s) and substituting $p_{js}^{\ell} = \mu_{ss'}^{\ell} c_j$,

$$Q_j \;=\; \sum_{\ell,s} N_s^{\ell}\, \widehat{D}_{ss'}^{\ell} \left(\frac{\mu_{ss'}^{\ell} c_j}{P_{ss'}^{\ell}} \right)^{-\sigma_{s'}} \;=\; c_j^{-\sigma_{s'}} \underbrace{\sum_{\ell,s} N_s^{\ell}\, (\mu_{ss'}^{\ell})^{-\sigma_{s'}} (P_{ss'}^{\ell})^{\sigma_{s'}}\, \widehat{D}_{ss'}^{\ell}}_{\mathcal{D}_{s'}}. \tag{18}$$

Hence

$$c_j^{1-\sigma_{s'}} \;=\; \left(\frac{Q_j}{\mathcal{D}_{s'}} \right)^{\frac{\sigma_{s'}-1}{\sigma_{s'}}}. \tag{19}$$

Step 2: variable (marginal–price) revenue. Revenue at the allocative margin from partition (ℓ, s) is

$$R_j^{\mathrm{lin};\ell,s} \;=\; \int p_{js}^{\ell} m_{ij}\, d\nu_{\ell s}(i) \;=\; N_s^{\ell}\, \widehat{D}_{ss'}^{\ell}\, (\mu_{ss'}^{\ell} c_j)^{1-\sigma_{s'}} (P_{ss'}^{\ell})^{\sigma_{s'}-1}.$$

Summing across partitions and using (19),

$$R_j^{\mathrm{lin}} \;=\; \left(\frac{Q_j}{\mathcal{D}_{s'}} \right)^{\frac{\sigma_{s'}-1}{\sigma_{s'}}} \underbrace{\sum_{\ell,s} N_s^{\ell}\, (\mu_{ss'}^{\ell})^{1-\sigma_{s'}} (P_{ss'}^{\ell})^{\sigma_{s'}-1}\, \widehat{D}_{ss'}^{\ell}}_{\mathcal{S}_{s'}}. \tag{20}$$

Step 3: flat–fee revenue. For partition (ℓ, s), the fee for the lowest buyer type $\underline{z}_s^{\ell}$ satisfies (cf. Proposition)

$$F_{js}^{\ell}(\underline{z}_s^{\ell}) \;=\; \frac{1}{\sigma_{s'}-1}\, P_{ss'}^{\ell}\, M_{is'}(\underline{z}_s^{\ell}) \left(\frac{p_{js}^{\ell}}{P_{ss'}^{\ell}} \right)^{1-\sigma_{s'}}.$$

Aggregating over buyers in (ℓ, s) yields

$$R_j^{\mathrm{fee};\ell,s} \;=\; N_s^{\ell}\, \frac{P_{ss'}^{\ell} M_{is'}(\underline{z}_s^{\ell})}{\sigma_{s'}-1}\, (\mu_{ss'}^{\ell} c_j)^{1-\sigma_{s'}} (P_{ss'}^{\ell})^{\sigma_{s'}-1}.$$

Summing across partitions and using (19),

$$R_j^{\mathrm{fee}} \;=\; \left(\frac{Q_j}{\mathcal{D}_{s'}} \right)^{\frac{\sigma_{s'}-1}{\sigma_{s'}}} \underbrace{\sum_{\ell,s} N_s^{\ell}\, (\mu_{ss'}^{\ell})^{1-\sigma_{s'}} (P_{ss'}^{\ell})^{\sigma_{s'}-1}\, \frac{P_{ss'}^{\ell} M_{is'}(\underline{z}_s^{\ell})}{\sigma_{s'}-1}}_{\mathcal{F}_{s'}}. \tag{21}$$

Step 4: revenue representation (homotheticity) and closed–form scale. Adding (20) and (21),

$$R_j \;=\; \left(\frac{Q_j}{\mathcal{D}_{s'}} \right)^{\frac{\sigma_{s'}-1}{\sigma_{s'}}} (\mathcal{S}_{s'} + \mathcal{F}_{s'}) \;=\; A_{s'}^{u}\, Q_j^{(\sigma_{s'}-1)/\sigma_{s'}}, \tag{22}$$

with

$$A^u_{s'} \;=\; \mathcal{D}_{s'}^{-\frac{\sigma_{s'}-1}{\sigma_{s'}}}\,(\mathcal{S}_{s'}+\mathcal{F}_{s'}),$$

$$\begin{aligned}
\mathcal{D}_{s'} &\equiv \sum_{\ell,s} N_s^\ell\,(\mu^\ell_{ss'})^{-\sigma_{s'}}(P^\ell_{ss'})^{\sigma_{s'}}\,\widehat{D}^\ell_{ss'},\\
\mathcal{S}_{s'} &\equiv \sum_{\ell,s} N_s^\ell\,(\mu^\ell_{ss'})^{1-\sigma_{s'}}(P^\ell_{ss'})^{\sigma_{s'}-1}\,\widehat{D}^\ell_{ss'},\\
\mathcal{F}_{s'} &\equiv \sum_{\ell,s} N_s^\ell\,(\mu^\ell_{ss'})^{1-\sigma_{s'}}(P^\ell_{ss'})^{\sigma_{s'}-1}\frac{P^\ell_{ss'}M_{is'}(\underline{z}^\ell_s)}{\sigma_{s'}-1}.
\end{aligned} \tag{23}$$

Thus, upstream revenue is homogeneous of degree $\psi^u_{s'} = (\sigma_{s'}-1)/\sigma_{s'}$ in own output, with shifter $A^u_{s'}$ comprising a variable (marginal–price) component $\mathcal{S}_{s'}$ and a flat–fee component $\mathcal{F}_{s'}$, both attenuated by the effective demand index $\mathcal{D}_{s'}$.

C.3 Kuhn–Tucker implementation of marginal–cost pricing

Following the logic of Theorem 1 in Baqaee and Farhi (2020a), consider a social planner who chooses final consumption C, final demands $\{y_j\}$, outputs $\{Q_j\}$, intermediate allocations $\{m_{ij}\}$, and masses of active producers $\{N_j\}$ to maximize $U(C)$ subject to (i) material balance for each variety j, (ii) per–producer feasibility with mass N_i of active buyers/producers i, and (iii) entry costs $\mathcal{K}_i$ paid in units of the final good:

$$Q_j - y_j - \sum_i N_i\, m_{ij} = 0, \qquad Q_i \le N_i\,\mathcal{F}_i(m_{i\cdot}), \qquad C \ge \sum_i \mathcal{K}_i N_i, \qquad \sum_j y_j = C.$$

Form the Lagrangian with Kuhn–Tucker multipliers v_j (material balance for good j), η_i (feasibility for producer i), and ϑ (final–good/entry resource):

$$\mathcal{L} = U(C) + \sum_j v_j\left(Q_j - y_j - \sum_i N_i m_{ij}\right) + \sum_i \eta_i(N_i\mathcal{F}_i(m_{i\cdot}) - Q_i) + \vartheta\left(C - \sum_i \mathcal{K}_i N_i\right).$$

FOCs and complementary slackness:

$$\frac{\partial \mathcal{L}}{\partial m_{ij}}: \quad -v_j N_i + \eta_i N_i \frac{\partial \mathcal{F}_i}{\partial m_{ij}} = 0 \;\Rightarrow\; v_j = \eta_i \frac{\partial \mathcal{F}_i}{\partial m_{ij}} \quad (\forall i,j).$$

$$\frac{\partial \mathcal{L}}{\partial Q_j}: \quad v_j - \eta_j = 0 \;\Rightarrow\; v_j = \eta_j \quad (\forall j).$$

$$\frac{\partial \mathcal{L}}{\partial y_j}: \quad -v_j + \lambda = 0 \;\Rightarrow\; v_j = \lambda \quad (\forall j), \quad \text{with } \lambda \text{ the multiplier on } \sum_j y_j = C.$$

$$\frac{\partial \mathcal{L}}{\partial C}: \quad U'(C) - \lambda + \vartheta = 0.$$

$$\frac{\partial \mathcal{L}}{\partial N_i}: \quad -\sum_j v_j m_{ij} + \eta_i \mathcal{F}_i(m_{i\cdot}) - \vartheta \mathcal{K}_i \le 0, \quad N_i \ge 0, \quad N_i\Big(-\sum_j v_j m_{ij} + \eta_i \mathcal{F}_i - \vartheta \mathcal{K}_i\Big) = 0.$$

Using $v_j = \eta_j$, the input FOCs become

$$\eta_j \;=\; \eta_i \,\frac{\partial \mathcal{F}_i}{\partial m_{ij}} \quad (\forall i, j),$$

which are the planner's cost–minimizing conditions: effective prices (proportional to η_j) equal marginal costs along every input link. The entry condition states that, at the optimum, the surplus created by an additional firm i (valued at $\eta_i \mathcal{F}_i - \sum_j \eta_j m_{ij}$) equals the entry cost valued at $\vartheta \mathcal{K}_i$ whenever $N_i > 0$.

Decentralized implementation with markups, rebates, and entry. Suppose in the decentralized economy each seller j charges a constant markup $\mu_j > 1$ on intermediate sales, and retailers in buyer sector s charge $\mu_s^r > 1$ to final consumers. Introduce ad valorem rebates on purchases so buyers face effective marginal prices

$$\tilde{p}_{ij} = (1 - t_j)\, p_{ij}, \qquad t_j \equiv 1 - \frac{1}{\mu_j}, \qquad \tilde{p}_s = (1 - t_s^r)\, p_s, \qquad t_s^r \equiv 1 - \frac{1}{\mu_s^r}.$$

Then $\tilde{p}_{ij} = c_j$ for intermediates and $\tilde{p}_s = \mathrm{MC}_s$ for retail. Firm i's cost minimization with $\tilde{p}_{ij}$ yields

$$\tilde{p}_{ij} = \widetilde{\eta}_i \,\frac{\partial \mathcal{F}_i}{\partial m_{ij}} \;\Rightarrow\; c_j = \widetilde{\eta}_i \,\frac{\partial \mathcal{F}_i}{\partial m_{ij}},$$

which coincides with the planner's FOCs after a common normalization of shadow values ($\widetilde{\eta_i} \propto \eta_i$). Hence the decentralized $\{m_{ij}\}$, $\{Q_j\}$, and C replicate the planner's allocation. Two–part tariffs' flat fees are infra–marginal and do not affect these FOCs.

Financing and entry. Let each active firm i pay a non–distortionary license $\mathcal{T}_i$ and let the government rebate $t_j p_{ij} m_{ij}$ and $t_s^r p_s y_s$ to buyers. Setting $\mathcal{T}_i = \vartheta \mathcal{K}_i$ ensures that the decentralized free–entry condition (operating profits net of input rebates minus the license equal zero) matches the planner's complementary slackness for N_i. Because licenses and flat fees are infra–marginal, they do not alter marginal conditions, while markups can remain strictly positive to fund entry costs. Government budget balance is achieved by choosing $\{\mathcal{T}_i\}_i$ to equal the present value of rebate outlays at the implemented allocation.

Therefore, the planner's allocation can be implemented even when firms charge markups: purchase–side rebates neutralize marginal wedges so that buyers face marginal costs in all nests, and entry costs are financed by non–distortionary fixed charges, preserving the planner's first–order conditions and entry margins.

C.4 Seller–identity invariance of the total unit markup

By the CES share rule within seller sector s', buyer i's demand for seller j's variety is

$$m_{ij} \;=\; M_{is'}\left(\frac{p^{\ell}_{js}}{P^{\ell}_{ss'}}\right)^{-\sigma_{s'}}.$$

From the optimal tariff characterization, the flat fee charged to buyer i in partition (ℓ, s) by seller $j \in \mathcal{U}_{s'}$ (with $\sigma_{s'} > 1$) is

$$F^{\ell}_{js} \;=\; \frac{1}{\sigma_{s'}-1}\,\tau_{is'}(\underline{z}^{\ell}_{s})\left(M_{is'}(\underline{z}^{\ell}_{s})\right)^{\frac{\sigma_{s'}-1}{\sigma_{s'}}}\left(\frac{p^{\ell}_{js}}{P^{\ell}_{ss'}}\right)^{1-\sigma_{s'}}, \qquad \tau_{is'} \equiv P^{\ell}_{ss'}\,M^{1/\sigma_{s'}}_{is'}.$$

Divide the fee by quantity using the share rule:

$$\frac{F^{\ell}_{js}}{m_{ij}} = \frac{p^{\ell}_{js}}{P^{\ell}_{ss'}}\cdot\frac{\frac{1}{\sigma_{s'}-1}\,\tau_{is'}(\underline{z}^{\ell}_{s})\left(M_{is'}(\underline{z}^{\ell}_{s})\right)^{\frac{\sigma_{s'}-1}{\sigma_{s'}}}}{M_{is'}}.$$

Total unit price is $\frac{T_{ij}}{m_{ij}} = p^{\ell}_{js} + \frac{F^{\ell}_{js}}{m_{ij}}$. Using $p^{\ell}_{js} = \mu^{\ell}_{ss'}\,c_j$ and dividing by c_j yields

$$\frac{\frac{T_{ij}}{m_{ij}}}{c_j} = \mu^{\ell}_{ss'}\left[1+\frac{1}{P^{\ell}_{ss'}}\cdot\frac{\frac{1}{\sigma_{s'}-1}\,\tau_{is'}(\underline{z}^{\ell}_{s})\left(M_{is'}(\underline{z}^{\ell}_{s})\right)^{\frac{\sigma_{s'}-1}{\sigma_{s'}}}}{M_{is'}}\right] \;\equiv\; \mu^{\ell}_{ss'}\Big(1+\chi^{\ell}_{ss'}(i)\Big).$$

The term $\chi^{\ell}_{ss'}(i)$ depends only on buyer–side objects within (ℓ, s) and on the sectoral index $P^{\ell}_{ss'}$, but not on seller j. Therefore the total unit markup is invariant to the seller's identity within a given buyer partition.

C.5 Profit Functions Results

Profits of the Lowest Type. For all retail sectors s, equilibrium profits of the lowest–productivity firm $i = (r, s, \underline{z}^{r}_{s})$ are not necessarily zero, nor necessarily positive. In particular,

$$\Pi(\underline{z}^{r}_{s}) > 0 \quad\Longleftrightarrow\quad \frac{1}{(1-\alpha^{r}_{s})(\varphi_s-1)} \;\geq\; \zeta_s, \qquad \zeta_s := \sum_{s'\in\mathcal{S}}\frac{\theta^{r}_{ss'}}{\sigma_{s'}-1}.$$

Consider a retail firm $i = (r, s, z_i)$ with $z_i = \underline{z}^{r}_{s}$. Denote by *variable profits* the component net of flat fees. Since retailers charge a constant markup $\varphi_s/(\varphi_s - 1)$, variable profits equal a fixed share

of revenue:

$$\text{VarProf}(\underline{z}_s^r) \;=\; \frac{1}{\varphi_s}\,\text{Revenue}(\underline{z}_s^r).$$

Total profits subtract the flat fees paid to upstream suppliers,

$$\Pi(\underline{z}_s^r) \;=\; \text{VarProf}(\underline{z}_s^r) \;-\; \sum_{s'\in\mathcal{S}} F_{ss'}^r\, N_{s'}^u.$$

Substituting yields

$$\Pi(\underline{z}_s^r) \;=\; \frac{\text{Revenue}(\underline{z}_s^r)}{\varphi_s}\Big[1-(1-\alpha_s^r)(\varphi_s-1)\,\zeta_s\Big].$$

Note that ζ_s is a weighted average of $\frac{1}{\sigma_{s'}-1}$ across seller sectors, with weights given by the Cobb–Douglas elasticities $\theta_{ss'}^r$, which satisfy $\sum_{s'}\theta_{ss'}^r = 1$. Thus, more important inputs in production receive greater weight in ζ_s.

Average profits. It is useful to restate how revenue scales with productivity. For retail and upstream firms we have

$$\text{Revenue}(z_s^r) = \left(\frac{z_s^r}{\widetilde{z}_s^r}\right)^{\varphi_s-1}\text{Revenue}(\underline{z}_s^r), \qquad \text{Revenue}(z_{s'}^u) = \left(\frac{z_{s'}^u}{\widetilde{z}_{s'}^u}\right)^{\sigma_{s'}-1}\text{Revenue}(\underline{z}_{s'}^u).$$

This scaling implies that average profits can be expressed in closed form.

For retailers,

$$\begin{aligned}\mathbb{E}\left[\Pi_s^r\right] &= \text{Revenue}(\widetilde{z}_s^r) - \sum_{s'\in\mathcal{S}} F_{ss'}^r\, N_{s'}^u \\ &= \frac{\text{Revenue}(\underline{z}_s^r)}{\varphi_s}\left[\left(\frac{\widetilde{z}_s^r}{\underline{z}_s^r}\right)^{\varphi_s-1} - (1-\alpha_s^r)(\varphi_s-1)\sum_{s'\in\mathcal{S}}\frac{\theta_{ss'}^r}{\sigma_{s'}-1}\right].\end{aligned}$$

For $j\in\mathcal{U}_{s'}$, expected profits decompose into (i) variable profits from the allocative margin, (ii) flat fees collected from buyers, and (iii) flat fees paid upstream:

$$\mathbb{E}\left[\Pi_j^u\right] = \underbrace{\sum_{\ell\in\{u,r\}}\sum_{s\in\mathcal{S}}\int_{i\in\mathcal{F}_{\ell s}}\left(p_{js}^\ell - c_j\right)m_{ij}\,d\nu_{\ell s}(i)}_{\text{variable profits from allocative margin}} + \underbrace{\sum_{\ell\in\{u,r\}}\sum_{s\in\mathcal{S}}\int_{i\in\mathcal{F}_{\ell s}} F_{js}^\ell(i)\,d\nu_{\ell s}(i)}_{\text{flat–fee revenue collected}} - \underbrace{\sum_{t\in\mathcal{S}}\int_{h\in\mathcal{U}_t} F_{hs'}^u(j)\,d\nu_t(h)}_{\text{flat–fee payments}}.$$

Here p_{js}^ℓ is the allocative (marginal) price charged by seller j to buyers in partition (ℓ,s), c_j is seller j's marginal cost, m_{ij} is buyer i's quantity purchased from j, $F_{js}^\ell(i)$ is the flat fee i pays to j, $\mathcal{F}_{\ell s}$ is the set of active buyers of type ℓ in sector s with measure $\nu_{\ell s}$, and the last term aggregates the flat fees $F_{hs'}^u(j)$ that j pays to its own upstream suppliers $h\in\mathcal{U}_t$.

Average upstream profits depend only on sectoral labor allocation. Fix technology and demand primitives $\{\alpha, \theta, \sigma\}$ and the productivity distributions (so that $\{\underline{z}, \widetilde{z}\}$ are fixed). Then the expected profit of the average upstream firm in sector s' depends only on sectoral labor allocation according to:

$$\mathbb{E}\left[\Pi^u_{s'}\right] = \frac{1}{N^u_{s'}}\left[\sum_{s\in\mathcal{S}} w\, L^r_s\, \Lambda^r_{ss'} \;+\; \sum_{t\in\mathcal{S}} w\, L^u_t\, \Lambda^u_{ts'}\right] \;-\; w\, l^u_{s'} \sum_{t\in\mathcal{S}} \Lambda^u_{s't},$$

where $L^\ell_s = l^\ell_s(\widetilde{z}^\ell_s)\, N^\ell_s$ is total labor used in sector (ℓ, s), and $l^\ell_s(\widetilde{z}^\ell_s)$ denotes labor of the average variety (productivity $\widetilde{z}^\ell_s$). The coefficients are

$$\Lambda^r_{ss'} \;=\; \frac{(1-\alpha^r_s)\,\theta^r_{ss'}}{\alpha^r_s}\left(\frac{1}{\sigma_{s'}-1}\left(\frac{\frac{z^r_s}{\widetilde{z}^r_s}}{\widetilde{\widetilde{z}}^r_s}\right)^{\varphi_s-1}+1\right), \qquad \Lambda^u_{s't} \;=\; \frac{(1-\alpha^u_{s'})\,\theta^u_{s't}}{\alpha^u_{s'}}\left(1+\frac{1}{\sigma_t-1}\left(\frac{\frac{z^u_{s'}}{\widetilde{z}^u_{s'}}}{\widetilde{\widetilde{z}}^u_{s'}}\right)^{\sigma_{s'}-1}\right).$$

Hence, conditional on primitives, $\mathbb{E}\Pi^u_{s'}$ varies solely with the sectoral labor aggregates $\{w\, L^\ell_s\}$ and own $w\, l^u_{s'}$[31].

C.6 Welfare decomposition

(1) Normalization and numeraire. We work in steady state and normalize the final consumer labor endowment to one. Let $\delta \in (0,1)$ be the per-period exit probability and m the mass of entrants. The law of motion implies $N = m/(1-\delta)$ at the sector level and in aggregate. Free entry equates expected discounted profits to entry costs in wage units, so with entry cost c_e,

$$\frac{\mathbb{E}[\pi]}{1-\delta} = w\, c_e \quad\Rightarrow\quad \mathbb{E}[\pi] = (1-\delta)\, w\, c_e.$$

Aggregate operating profits are $\Pi = N\,\mathbb{E}[\pi] = \frac{m}{1-\delta}\cdot(1-\delta)\, w\, c_e = m\, w\, c_e$. Nominal income is $Y = w\, L_{\text{prod}} + \Pi = w\Big(L_{\text{prod}} + m\, c_e\Big)$. With unit labor endowment, $L_{\text{prod}} + m\, c_e = 1$, hence $Y = w$. Because indirect utility is homogeneous of degree zero in prices and income, measuring utility in wage units yields $W = \frac{Y}{P_Y} = \frac{w}{P_Y} = \frac{1}{P_Y}$.

(2) Objects and dimensions. Let $b := (\theta_s)_{s\in\mathcal{S}} \in \mathbb{R}^{1\times|\mathcal{S}|}$ be final-demand shares (row), and let the cost-based input–output blocks be

$$\Omega^{ru},\ \Omega^{uu} \in \mathbb{R}^{|\mathcal{S}|\times|\mathcal{S}|}, \qquad \Omega^{ru}_{ss'} = (1-\alpha^r_s)\,\theta^r_{ss'}, \quad \Omega^{uu}_{ss'} = (1-\alpha^u_s)\,\theta^u_{ss'}.$$

[31] Sketch. Start from the upstream profits decomposition . Use (i) p–P CES shares and the identity $(Q^u_{s'}/D_{s'})^{(\sigma_{s'}-1)/\sigma_{s'}} = 1/N^u_{s'}$, (ii) the flat–fee formula $F^\ell_{ss'} = \frac{P^\ell_{ss'} M^\ell_{ss'}(z^\ell_s)}{N^u_{s'}(\sigma_{s'}-1)}$, and (iii) cost minimization $P^\ell_{ss'} M^\ell_{ss'} = w\, l^\ell_s \frac{(1-\alpha^\ell_s)\theta^\ell_{ss'}}{\alpha^\ell_s}$ to aggregate across buyers and suppliers. Collecting terms yields the stated affine function of $\{w\, L^\ell_s\}$ and $w\, l^u_{s'}$ with coefficients Λ.

The upstream Leontief inverse is

$$\Psi^{uu} := (I - \Omega^{uu})^{-1} = \sum_{n\geq 0} (\Omega^{uu})^n \quad \text{(well-defined if } \rho(\Omega^{uu}) < 1\text{)}.$$

With a separable final-good aggregator across retail sectors, the consumer–retail map is the identity, so b already captures direct exposure to retail prices.

Exposure (Domar-type) scalars. Define sectoral exposure *numbers*, for each $s \in \mathcal{S}$,

$$\tilde{\lambda}_s^{cr} := b_s, \qquad \tilde{\lambda}_s^{ru} := \sum_v \tilde{\lambda}_v^{cr}\, \Omega_{vs}^{ru}, \qquad \tilde{\lambda}_s^{uu} := \sum_v \tilde{\lambda}_v^{ru}\, \Psi_{vs}^{uu},$$

or, in vector form,

$$\tilde{\lambda}^{cr} = b, \qquad \tilde{\lambda}^{ru} = b\,\Omega^{ru}, \qquad \tilde{\lambda}^{uu} = b\,\Omega^{ru}\,\Psi^{uu}.$$

These mirror the Baqaee–Farhi cost-based Domar weights $b\Psi$, adapted to the consumer → retail → upstream layering.

(3) Sectoral indices and the upstream recursion. For retail sector s,

$$\log P_s = \log \mu_s^r + \sum_{s'} \Omega_{ss'}^{ru} \Big(\log \mu_{s's}^r + \log C_{s'} \Big) - \frac{1}{\varphi_s - 1} \log N_s^r + \text{const} + \log \mathcal{V}_s, \tag{24}$$

where $P_{s's}^r = \mu_{s's}^r C_{s'}$. For upstream sector s',

$$\log C_{s'} = \sum_v \Omega_{s'v}^{uu} \Big(\log \mu_{vs'}^u + \log C_v \Big) - \frac{1}{\sigma_{s'} - 1} \log N_{s'}^u + \text{const} + \log \mathcal{V}_{s'}^u. \tag{25}$$

Stacking (25) across s' gives the linear recursion

$$\log C^u = \Omega^{uu} \log C^u + \log \mu^{uu} - \frac{\log N^u}{\sigma - 1} + \log \mathcal{V}^u + \text{const}, \tag{26}$$

so for differences between two equilibria (technology and numeraire drop out),

$$\Delta \log C^u = \Psi^{uu} \left(\Delta \log \mu^{uu} - \frac{\Delta \log N^u}{\sigma - 1} + \Delta \log \mathcal{V}^u \right). \tag{27}$$

(4) From sectoral indices to the final index. Taking differences in (24) and substituting (27) yields, for each s,

$$\Delta \log P_s = \Delta \log \mu_s^r + \sum_{s'} \Omega_{ss'}^{ru}\, \Delta \log \mu_{s's}^r + \sum_{s'} \Omega_{ss'}^{ru}\, \Delta \log C_{s'} - \frac{1}{\varphi_s - 1} \Delta \log N_s^r + \Delta \log \mathcal{V}_s,$$

with

$$\Delta \log C = \Psi^{uu} \left(\Delta \log \mu^{uu} - \frac{\Delta \log N^u}{\sigma - 1} + \Delta \log \mathcal{V}^u \right).$$

Aggregate,

$$\Delta \log P_Y = b \, \Delta \log \mu^r + b \, \Omega^{ru} \underbrace{\mathrm{Agg}_r(\Delta \log \mu^r_{s's})}_{\text{retail–upstream wedges}} + b \, \Omega^{ru} \, \Psi^{uu} \left(\Delta \log \mu^{uu} - \frac{\Delta \log N^u}{\sigma - 1} + \Delta \log \mathcal{V}^u \right) \quad (28)$$

$$- b \, \frac{\Delta \log N^r}{\varphi - 1} + b \, \Delta \log \mathcal{V}, \quad (29)$$

where $\mathrm{Agg}_r(\cdot)$ denotes the linear aggregation of buyer-specific retail–upstream wedges into a sectoral vector (defined next).

(5) Aggregating buyer-specific wedges and exposure maps. Define sectoral wedge vectors by linear aggregation with Ω-weights:

$$\left[\log \mu^r \right]_{s'} := \frac{1}{\bar{b}_{s'}} \sum_s b_s \, \Omega^{ru}_{ss'} \, \log \mu^r_{s's}, \qquad \bar{b} := b \, \Omega^{ru} \in \mathbb{R}^{1 \times |\mathcal{S}|},$$

and

$$\left[\log \mu^{uu} \right]_{s'} := \sum_v \Omega^{uu}_{s'v} \, \log \mu^u_{vs'}.$$

Then the linear maps in (28) reduce to contractions with the exposure vectors:

$$b \, \Omega^{ru} \; \mathrm{Agg}_r(\Delta \log \mu^r_{s's}) = (b \, \Omega^{ru}) \, \Delta \log \mu^r = \tilde{\lambda}^{ru} \, \Delta \log \mu^r, \qquad b \, \Omega^{ru} \, \Psi^{uu} \, \Delta \log \mu^{uu} = \tilde{\lambda}^{uu} \, \Delta \log \mu^{uu}.$$

Similarly,

$$b \, \Delta \log \mu^r = \sum_s \tilde{\lambda}^{cr}_s \, \Delta \log \mu^r_s, \quad b \, \frac{\Delta \log N^r}{\varphi - 1} = \sum_s \frac{\tilde{\lambda}^{cr}_s}{\varphi_s - 1} \, \Delta \log N^r_s, \quad b \, \Omega^{ru} \, \Psi^{uu} \, \frac{\Delta \log N^u}{\sigma - 1} = \tilde{\lambda}^{uu} \left(\frac{\Delta \log N^u}{\sigma - 1} \right).$$

(6) Conclusion (and selection). Using $\Delta \log W = -\Delta \log P_Y$ in (28) and the identities above gives

$$\Delta \log W = \underbrace{- \sum_s \tilde{\lambda}^{cr}_s \, \Delta \log \mu^r_s - \tilde{\lambda}^{ru} \, \Delta \log \mu^r - \tilde{\lambda}^{uu} \, \Delta \log \mu^{uu}}_{\text{intensive (markups)}} + \underbrace{\sum_s \frac{\tilde{\lambda}^{cr}_s}{\varphi_s - 1} \, \Delta \log N^r_s + \tilde{\lambda}^{uu} \left(\frac{\Delta \log N^u}{\sigma - 1} \right)}_{\text{extensive (firm masses)}}$$

$$- \underbrace{\left(b \, \Delta \log \mathcal{V} + \tilde{\lambda}^{uu} \, \Delta \log \mathcal{V}^u \right)}_{\text{selection}}. \quad (30)$$

If the composition of active varieties is invariant (e.g., Pareto tails with unchanged truncation), then $\Delta \log \mathcal{V} = \Delta \log \mathcal{V}^u = 0$ and the selection term vanishes, yielding Proposition 2.

C.7 Equilibrium existence and uniqueness

Roadmap. The decentralized equilibrium with two–part tariffs is pinned in six linked steps. S1 fixes the composition of retail labor from final demand. S2 maps retail labor into upstream labor through a linear network that depends on cost shares and buyer–specific μ markups. S3 stacks upstream free entry in labor units, yielding a linear relation between sectoral labor and upstream entry flows. S4 composes S2 and S3 and adds retail free entry, so entry flows on both layers are linear in retail labor. S5 imposes aggregate labor clearing on the retail ray, reducing the problem to one scalar t with a unique solution. S6 solves the price–cost block (log indices and wage) from linear CES/Cobb–Douglas relations; the coefficient matrix is a contraction. Under S1-S6, we can show equilibrium existances and uniqueness, where, quantities, masses, and prices are all uniquely determined.

Sectors are indexed by $s, m \in \{1, \ldots, S\}$. Layer $\ell \in \{u, r\}$ denotes upstream or retail. Labor shares $\alpha_s^\ell \in (0,1)$; materials shares sum to one by buyer: $\sum_m \theta_{u,m,\cdot} = 1$ for upstream buyers and $\sum_m \theta_{r,\cdot,m} = 1$ for retail buyers. Elasticities $\sigma_m > 1$ (upstream) and $\varphi_s > 1$ (retail). Final demand across retail sectors is Cobb–Douglas with weights $\boldsymbol{\theta}_c = (\theta_{c,1}, \ldots, \theta_{c,S})$, $\mathbf{1}^\top \boldsymbol{\theta}_c = 1$. Buyer–specific per–unit markups are denoted by μ: $\mu_{s,m}^{u \to u}$ (upstream buyer s from upstream seller m), $\mu_{s,m}^{u \to r}$ (retail buyer s from upstream seller m), and μ_s^r (retail to final consumer in sector s). Entry costs $c_{e,s}^\ell > 0$; exit rates $\delta_s^\ell \in [0,1)$. Masses N_s^ℓ, entry flows $e_s^\ell = (1 - \delta_s^\ell) N_s^\ell$. Stack vectors by sector: $\mathbf{L}_u, \mathbf{L}_r, \mathbf{e}_u, \mathbf{e}_r \in \mathbb{R}_+^S$. Lower case denotes logs (e.g. $w = \ln W$, $p = \ln P$).

Equilibrium conditions

1. Retail free entry: $\mathbb{E}[\Pi_s^r] = W c_{e,s}^r (1 - \delta_s^r)$ for all s.

2. Upstream free entry: $\mathbb{E}[\Pi_m^u] = W c_{e,m}^u (1 - \delta_m^u)$ for all m.

3. Average firm output meets demand within sector:

$$\widetilde{z}_{r,s}\, y_{r,s}(\widetilde{z}_{r,s}) = Y_{c,\cdot,s}\, N_{r,s}^{\frac{\varphi_s}{1-\varphi_s}}, \qquad \widetilde{z}_{u,m}\, y_{u,m}(\widetilde{z}_{u,m}) = D_{u,m}\, N_{u,m}^{\frac{\sigma_m}{1-\sigma_m}}.$$

4. CES unit–price indices for any buyer (b, s) from upstream seller m:

$$p_{b,s,m} = \ln \mu_{s,m}^{u \to b} + mc_{u,m} + \frac{1}{1 - \sigma_m}\, n_{u,m}.$$

   Upstream marginal cost: $mc_{u,m} = (\ln \Theta_{u,m} - \ln \widetilde{z}_{u,m}) + \alpha_m^u w + (1 - \alpha_m^u) \sum_j \theta_{u,m,j}\, p_{u,m,j}$. Retail marginal cost: $mc_{r,s} = (\ln \Theta_{r,s} - \ln \widetilde{z}_{r,s}) + \alpha_s^r w + (1 - \alpha_s^r) \sum_m \theta_{r,s,m}\, p_{r,s,m}$. Final–good index: $p_{c,s} =$

$\ln \mu_s^r + mc_{r,s} + \frac{1}{1-\varphi_s} n_{r,s}$, $\sum_s \theta_{c,s} p_{c,s} = 0$.

5. Labor market clearing: $1 = \sum_s L_s^r + \sum_m L_m^u + \sum_s e_s^r + \sum_m e_m^u$.

S1. Final demand pins the retail composition (ray)

Cobb–Douglas final demand implies sectoral retail revenue shares equal preference weights. With CRS, $WL_s^r = \alpha_s^r \text{Revenue}_s^r$, so relative retail labor is fixed:

$$\bar{\mathbf{L}}_r \propto \left(\frac{\theta_{c,s}}{\alpha_s^r}\right)_{s=1}^{S} \gg 0, \qquad \mathbf{L}_r = t\,\bar{\mathbf{L}}_r, \quad t > 0.$$

Only the scalar t remains to be determined by aggregate labor clearing.

S2. Sectoral labor mapping with buyer–specific markups

At the sector level, CRS cost shares and CES demand under buyer–specific per–unit markups imply a linear system that maps retail labor into upstream labor and propagates upstream feedback:

$$\mathbf{L}_u = A\,\mathbf{L}_u + B\,\mathbf{L}_r,$$

with nonnegative $S \times S$ matrices

$$A_{s,m} = \alpha_s^u \frac{1-\alpha_m^u}{\alpha_m^u} \frac{\theta_{u,m,s}}{\mu_{m,s}^{u\to u}}, \qquad B_{s,r} = \alpha_s^u \frac{1-\alpha_r^r}{\alpha_r^r} \frac{\theta_{r,r,s}}{\mu_{r,s}^{u\to r}}.$$

Interpretation: a buyer in upstream sector s allocates the share $(1-\alpha_m^u)\theta_{u,m,s}$ of revenue to materials from upstream seller m, scaled by the buyer–specific markup faced on that link; the labor anchor α_s^u converts revenue to labor. Likewise for retail exposure B. Assume the spectral condition

$$\rho(A) < 1 \quad \text{(sufficient: each row sum of } A \text{ is } < 1),$$

so the Neumann series converges and the total upstream requirements per unit of retail labor are

$$\Psi := (I - A)^{-1} B \geq 0, \qquad \mathbf{L}_u = \Psi\,\mathbf{L}_r.$$

Higher per–unit markups weaken links (prices are higher), reducing the corresponding entries of A and B and, through Ψ, the upstream labor implied by a given $\mathbf{L}_r$.

S3. Upstream free entry in labor units (stacked)

Two–part tariffs imply that an upstream entrant's expected period profit equals a constant fraction of buyer spending (variable margin) plus net flat–fee transfers (received from buyers minus paid

to own suppliers). Dividing by W makes profits linear in buyer revenues expressed in labor units:

$$(G^{uu} - \Pi_u)\,\mathbf{L}_u \;+\; G^{ru}\,\mathbf{L}_r \;=\; C_u\,\mathbf{e}_u,$$

where $G^{uu}, G^{ru} \geq 0$ are flow–weight matrices determined by cost shares and the tariff schedule, $\Pi_u = \mathrm{diag}((G^{uu})^\top \mathbf{1})$ nets out intra–upstream transfers, $C_u = \mathrm{diag}(c^u_{e,s})$, and $\mathbf{e}_u$ stacks upstream entry flows. The left side converts sectoral labor into expected upstream profits by sector; the right side is entry cost (in labor units) times the entrant flow, as required by free entry.

S4. Entry maps as linear functions of retail labor

Compose S2 into S3:

$$\mathbf{e}_u = C_u^{-1}\Big[(G^{uu} - \Pi_u)\Psi + G^{ru}\Big]\,\mathbf{L}_r \;=\; \chi_u\,\mathbf{L}_r, \qquad \chi_u \geq 0.$$

Retail free entry implies $e^r_s = (1-\delta^r_s)N^r_s = (1-\delta^r_s)L^r_s/l^r_s$. Let $\Gamma_r = \mathrm{diag}((1-\delta^r_s)/l^r_s) \succ 0$. Then

$$\mathbf{e}_r = \Gamma_r\,\mathbf{L}_r.$$

Thus, once the retail composition $\bar{\mathbf{L}}_r$ is fixed by S1, both upstream and retail entry flows move linearly with the scale t along that ray.

S5. Labor clearing on the retail ray

Total labor equals production labor plus entry labor:

$$1 = \mathbf{1}^\top\Big(\mathbf{L}_r + \mathbf{L}_u + \mathbf{e}_r + \mathbf{e}_u\Big) = \mathbf{1}^\top\Big[\Big(I + \Psi + \Gamma_r + \chi_u\Big)\,\mathbf{L}_r\Big].$$

On the ray $\mathbf{L}_r = t\,\bar{\mathbf{L}}_r$, define

$$\Xi(t) := \mathbf{1}^\top\Big(I + \Psi + \Gamma_r + \chi_u\Big)\,(t\,\bar{\mathbf{L}}_r).$$

Because $(I + \Psi + \Gamma_r + \chi_u)\bar{\mathbf{L}}_r \gg 0$, Ξ is continuous, strictly increasing, $\Xi(0) = 0$, and $\Xi(t) \to \infty$ as $t \to \infty$. There exists a unique $t^\star > 0$ such that $\Xi(t^\star) = 1$. This pins

$$\mathbf{L}^\star_r = t^\star \bar{\mathbf{L}}_r, \qquad \mathbf{L}^\star_u = \Psi\,\mathbf{L}^\star_r, \qquad \mathbf{e}^\star_u = \chi_u\,\mathbf{L}^\star_r, \qquad \mathbf{e}^\star_r = \Gamma_r\,\mathbf{L}^\star_r,$$

and, hence, sectoral masses $N^\ell_s = e^\ell_s/(1-\delta^\ell_s)$.

S6. Price/cost block under wage normalization

The nominal wage is the numeraire. Write the nominal wage as ω and set $\omega = 1$, so its log is $w = \ln\omega = 0$. This step solves for sectoral price indices conditional on the masses from S1–S5, and

then computes welfare.

For each buyer layer–sector pair $(b,s) \in \{u,r\} \times \{1,\dots,S\}$ and upstream seller m, the CES unit price index satisfies

$$p_{b,s,m} \;=\; \ln \mu_{m,s}^{u\to b} \;+\; mc_{u,m} \;+\; \frac{1}{1-\sigma_m}\, n_{u,m}, \tag{P*}$$

where $p_{b,s,m} = \ln P_{b,s,m}$, $n_{u,m} = \ln N_m^u$, and $\mu_{m,s}^{u\to b}$ is the buyer–specific markup applied by upstream sector m when selling to buyer (b,s). Upstream marginal costs in logs are

$$mc_{u,m} \;=\; \left(\ln \Theta_{u,m} - \ln \widetilde{z}_{u,m}\right) \;+\; (1-\alpha_m^u) \sum_{j=1}^{S} \theta_{u,m,j}\, p_{u,m,j} \tag{MC–U*}$$

and retail marginal costs are

$$mc_{r,s} \;=\; \left(\ln \Theta_{r,s} - \ln \widetilde{z}_{r,s}\right) \;+\; (1-\alpha_s^r) \sum_{m=1}^{S} \theta_{r,s,m}\, p_{r,s,m} \tag{MC–R*}$$

Final–good price indices by retail sector are

$$p_{c,s} \;=\; \ln \mu_s^r \;+\; mc_{r,s} \;+\; \frac{1}{1-\varphi_s}\, n_{r,s}, \tag{PC*}$$

where $n_{r,s} = \ln N_s^r$ and μ_s^r is the retail–to–consumer markup in sector s. For notational economy define the technology–selection constants

$$\xi_{u,m} := \ln \Theta_{u,m} - \ln \widetilde{z}_{u,m}, \qquad \xi_{r,s} := \ln \Theta_{r,s} - \ln \widetilde{z}_{r,s}.$$

Collect upstream–to–upstream indices by seller into the vector $p_u^u \in \mathbb{R}^S$ and define the non-negative matrix

$$B_u \;:=\; \operatorname{diag}(1-\alpha^u)\,\Theta_u \;\in\; \mathbb{R}^{S\times S}, \qquad \|B_u\|_\infty = \max_m (1-\alpha_m^u) < 1.$$

Using (P*) and (MC–U*) in seller–stacked form,

$$(I - B_u)\, p_u^u \;=\; \ln \mu^{\,u\to u} \;+\; \xi_u \;+\; \frac{1}{1-\sigma}\, n_u, \qquad p_u^r \;=\; p_u^u \;+\; \ln\!\left(\mu^{\,u\to r} \oslash \mu^{\,u\to u}\right), \tag{U*}$$

where $\oslash$ denotes componentwise division and n_u, ξ_u, $\ln \mu^{\,u\to\ell} \in \mathbb{R}^S$ are the seller–indexed vectors. Because $\|B_u\|_\infty < 1$, $I - B_u$ is invertible and the upstream fixed point is unique.

Retail sector indices follow from (P*) and (MC–R*) as

$$p_r[s] \;=\; \ln \mu_s^r \;+\; \xi_{r,s} \;+\; (1-\alpha_s^r)\sum_{m=1}^{S} \theta_{r,s,m}\, p_u^r[m] \;+\; \frac{1}{1-\varphi_s}\, n_{r,s}. \tag{R*}$$

With the wage fixed at one, real income equals the inverse final price index. Writing the final–demand weights as $\boldsymbol{\theta}_c = (\theta_{c,1}, \ldots, \theta_{c,S})^\top$ and $\Omega_c := \boldsymbol{\theta}_c^\top$, welfare is

$$\mathcal{W} \;=\; -\,\Omega_c\, p_c \;=\; -\sum_{s=1}^{S} \theta_{c,s}\, p_{c,s}, \tag{W*}$$

with p_c obtained from (PC*) using p_u^u, p_u^r, and p_r computed above.

Existence and uniqueness of the price block under wage normalization. The upstream map $p_u^u \;\mapsto\; (I - B_u)^{-1}\Big(\ln \mu^{u\to u} + \xi_u + \frac{1}{1-\sigma} n_u\Big)$ is well defined and single–valued because $\|B_u\|_\infty < 1$ implies $\rho(B_u) < 1$ and hence $I - B_u$ is invertible. Given p_u^u, the transformation to p_u^r is an additive constant shift governed by buyer–specific markup gaps, therefore also unique. Retail indices are then affine in p_u^r, and final–good indices are affine in mc_r and n_r, so both are uniquely pinned. The entire price/cost block is linear in the unknown indices once masses n_u, n_r are taken from S1–S5, and the numeraire has already fixed the absolute price level by setting $\omega = 1$. No additional normalization is required, and uniqueness follows from the upstream contraction and forward substitution in the remaining equations.

Existence and uniqueness with alternative price regimes Nonlinear pricing with two–part tariffs retains buyer–specific per–unit markups $\mu_{m,s}^{u\to b}$ and permits flat fees that do not enter unit price indices. Steps S1–S5 are unchanged, since they rely on constant–returns shares and free entry in labor units. In S6, the only difference is in the constants of the upstream and retail equations through $\ln \mu$; the upstream coefficient matrix $I - B_u$ and the retail aggregation weights remain the same. Because $\|B_u\|_\infty < 1$ still holds, the upstream fixed point is unique and the rest of the system follows uniquely by forward substitution. Existence and uniqueness of the full equilibrium therefore carry over under nonlinear pricing with wage normalization.

Planner–implemented marginal–cost pricing sets effective per–unit markups to one, $\mu^{u\to u} \equiv \mu^{u\to r} \equiv 1$. The price block strictly simplifies: equation (U*) becomes $(I - B_u)\, p_u^u = \xi_u + \frac{1}{1-\sigma} n_u$ with the same B_u, and $p_u^r = p_u^u$. Retail indices follow from (R*) with $\ln \mu^r$ set to zero if the planner eliminates retail markups. The contraction property is preserved, so prices are uniquely pinned under $\omega = 1$. Regarding the quantity side, either the planner chooses masses directly by appropriate entry transfers, in which case S1–S5 run with masses treated as given, or the planner supports decentralized free entry with linear subsidies that preserve the linear mapping from labor to entry in

labor units. In both implementations the one–dimensional labor–clearing step and the linear price block continue to deliver a unique equilibrium under the same spectral condition on upstream feedback, now in a system with strictly simpler coefficients.

D Parameter Calibration and Estimation

D.1 Labor Output Elasticity α

In the Chilean firm balance sheet accounts data, we observe expenditures on labor (wL), capital services (rK), and intermediate materials (M). Under cost minimization, the model's labor (non–material) output elasticity coincides with the variable-cost share of the non–material bundle (labor + capital) in total variable cost. Since the model abstracts from capital as a separate input, we bundle labor and capital into a single "non–material" composite measured in the data as $wL + rK$, and define for firm i:

$$\alpha_i \;=\; 1 - \frac{\sum_j p_{ji} m_{ji}}{w_i L_i + r_i K_i + \sum_j p_{ji} m_{ji}}\,.$$

If upstream suppliers charge two–part tariffs, total payments satisfy $TC_i = F_i + VC_i$ with $VC_i := w_i L_i + r_i K_i + \sum_j p_{ji} m_{ji}$. The mapping above holds at the level of variable costs. Total–cost shares equal variable–cost shares scaled by $(1 - esc_i)$, where $esc_i := F_i/(F_i + VC_i)$ is the flat–fee expenditure share. For large buyers (high VC_i), esc_i is small, so total–cost and variable–cost shares are close; throughout we therefore use α_i as the labor (non–material) output elasticity.

We keep firms above the 75$^{\text{th}}$ percentile of annual revenue, winsorize α_i at the 1st and 99th percentiles, compute $\alpha_{s,\ell}$ by 6–digit sector s and firm type $\ell \in \{\text{Retailers}, \text{Upstream}\}$ separately by year (2005–2022), then average over time and aggregate to the model's 11 sectors.

We report α under two evaluation lenses that interpret observed unit prices differently. Under the nonlinear–pricing lens, we exploit that under two–part tariffs average unit prices converge to the marginal price as quantity rises; hence, for large buyers, observed unit prices closely approximate marginal (allocative) prices. Under the uniform-pricing lens, we re-parameterize treating observed unit prices as marginal prices that are invariant to quantity and common across buyers within seller–product–time cells; since uniform pricing implies no fixed fees ($F_i = 0$, total and variable costs coincide, so we estimate α on the full firm population. Table A5 presents sectoral means by firm type under both lenses.

Table A5: Labor (non–material) shares by sector and firm type: nonlinear vs. uniform lenses

	Nonlinear lens α		Uniform lens α	
Sector	Retailers	Upstream	Retailers	Upstream
Agriculture	0.43	0.41	0.53	0.50
Mining	0.25	0.32	0.38	0.43
Manufacturing	0.39	0.42	0.49	0.59
Utilities	0.37	0.58	0.53	0.51
Construction	0.48	0.42	0.63	0.51
Retail and Wholesale	0.37	0.31	0.50	0.50
Transport and ICTs	0.55	0.47	0.66	0.58
Financial Services	0.58	0.62	0.77	0.77
Real Estate Services	0.66	0.53	0.75	0.67
Business Services	0.62	0.65	0.76	0.69
Personal Services	0.71	0.57	0.74	0.62
Type mean	0.49	0.48	0.61	0.58

Notes: α is the non–material (labor + capital) variable–cost share. Sectoral means pool 2005–2022 firm–year estimates at 6–digit × firm type, aggregated to 11 sectors. The nonlinear lens relies on large–buyer moments so average unit prices approximate marginal prices under two–part tariffs; the uniform lens treats per–unit prices as marginal and common across buyers (quantity–invariant).

On average, the uniform-pricing lens yields higher α because it treats observed per-unit prices as marginal across buyers and weights smaller, more labor-intensive firms more heavily, whereas under the nonlinear-pricing lens large-buyer moments (with flat-fee dilution) lower the measured non-material share.

D.2 Input–output and Output Elasticities

We recover buyer–facing variable expenditure shares (weights) on upstream seller sectors from firm–to–firm transactions. For buyer firm i of type $\ell \in \{r, u\}$, let $\mathcal{U}_{s'}$ denote the set of upstream varieties in seller sector s'. Define the materials expenditure weight on sector s' as:

$$\theta^{\ell}_{is'} := \frac{\sum_{j \in \mathcal{U}_{s'}} p_{ij}\, m_{ij}}{\sum_{s''} \sum_{j \in \mathcal{U}_{s''}} p_{ij}\, m_{ij}}, \qquad \sum_{s'} \theta^{\ell}_{is'} = 1,$$

where p_{ij} is the buyer–facing unit price and m_{ij} the corresponding quantity. Under two–part tariffs, total payments satisfy $TC_i = F_i + VC_i$ with $VC_i := w_i L_i + r_i K_i + \sum_j p_{ij} m_{ij}$; the weights above are defined on variable materials expenditure. For large buyers (high VC_i), the flat–fee share F_i/TC_i is

small, so total–cost shares closely track variable–cost shares.

Construction follows the logic used for α. For the nonlinear–pricing lens, we compute firm–level weights θ_{is}^{ℓ}, separately for retailers ($\ell = r$) and upstream buyers ($\ell = u$), retain firms above the 75th revenue percentile each year, aggregate from 6–digit industries to the buyer's 1–digit sector within year by simple averaging, and then average over 2005–2022. For the uniform–pricing lens, we repeat the construction on the full firm population, treating per–unit prices as marginal and common across buyers within seller–product–time cells. Rows sum to one up to rounding. The four matrices below report retailers and upstream firms as buyers under each lens.

Table A6: Input–output weights by Retailers as buyers (nonlinear–pricing lens)

Buyer / Seller	Agr.	Min.	Man.	Uti.	Cons.	R. & W.	T. & ICTs	F. Serv.	RE. Serv.	B. Serv.	P. Serv.
Agriculture	0.21	0.00	0.39	0.02	0.02	0.26	0.07	0.02	0.00	0.02	0.00
Mining	0.00	0.04	0.18	0.02	0.05	0.48	0.06	0.01	0.00	0.15	0.00
Manufacturing	0.10	0.01	0.42	0.02	0.01	0.32	0.08	0.01	0.00	0.04	0.00
Utilities	0.05	0.02	0.35	0.02	0.02	0.17	0.09	0.03	0.00	0.25	0.00
Construction	0.07	0.00	0.23	0.01	0.07	0.25	0.14	0.02	0.00	0.20	0.00
Retail and Wholesale	0.12	0.01	0.37	0.01	0.01	0.30	0.08	0.03	0.00	0.06	0.00
Transport and ICTs	0.05	0.01	0.26	0.02	0.02	0.16	0.24	0.02	0.00	0.22	0.00
Financial Services	0.06	0.00	0.21	0.01	0.01	0.23	0.06	0.12	0.00	0.30	0.00
Real Estate Services	0.03	0.00	0.24	0.01	0.01	0.25	0.03	0.04	0.03	0.35	0.00
Business Services	0.05	0.00	0.16	0.01	0.01	0.25	0.06	0.04	0.00	0.42	0.00
Personal Services	0.05	0.00	0.27	0.01	0.01	0.19	0.06	0.06	0.00	0.35	0.00

Table A7: Input–output weights by Retailers as buyers (uniform–pricing lens)

Buyer / Seller	Agr.	Min.	Man.	Uti.	Cons.	R. & W.	T. & ICTs	F. Serv.	RE. Serv.	B. Serv.	P. Serv.
Agriculture	0.25	0.00	0.21	0.02	0.03	0.32	0.05	0.07	0.00	0.04	0.00
Mining	0.00	0.04	0.19	0.06	0.15	0.30	0.07	0.02	0.00	0.17	0.00
Manufacturing	0.13	0.02	0.35	0.02	0.03	0.25	0.11	0.03	0.00	0.06	0.00
Utilities	0.07	0.01	0.18	0.03	0.03	0.26	0.17	0.05	0.00	0.20	0.00
Construction	0.10	0.00	0.10	0.02	0.22	0.24	0.15	0.03	0.00	0.14	0.00
Retail and Wholesale	0.16	0.01	0.24	0.01	0.02	0.34	0.08	0.05	0.00	0.09	0.00
Transport and ICTs	0.07	0.01	0.14	0.02	0.03	0.24	0.19	0.04	0.00	0.26	0.00
Financial Services	0.08	0.00	0.12	0.01	0.01	0.22	0.06	0.15	0.01	0.33	0.00
Real Estate Services	0.03	0.00	0.12	0.01	0.02	0.30	0.04	0.06	0.05	0.37	0.00
Business Services	0.07	0.00	0.13	0.01	0.01	0.22	0.09	0.06	0.00	0.41	0.00
Personal Services	0.07	0.00	0.17	0.02	0.02	0.25	0.07	0.08	0.00	0.33	0.01

Table A8: Input–output weights by Upstream firms as buyers (nonlinear–pricing lens)

Buyer / Seller	Agr.	Min.	Man.	Uti.	Cons.	R. & W.	T. & ICTs	F. Serv.	RE. Serv.	B. Serv.	P. Serv.
Agriculture	0.26	0.00	0.12	0.02	0.04	0.29	0.10	0.06	0.00	0.10	0.00
Mining	0.01	0.07	0.39	0.05	0.06	0.13	0.11	0.03	0.00	0.15	0.00
Manufacturing	0.08	0.02	0.49	0.03	0.02	0.15	0.09	0.02	0.00	0.10	0.00
Utilities	0.06	0.02	0.18	0.07	0.03	0.18	0.15	0.04	0.00	0.27	0.00
Construction	0.07	0.00	0.14	0.03	0.30	0.18	0.12	0.03	0.00	0.13	0.00
Retail and Wholesale	0.12	0.01	0.27	0.01	0.02	0.38	0.07	0.03	0.00	0.10	0.00
Transport and ICTs	0.06	0.02	0.14	0.02	0.04	0.21	0.22	0.03	0.00	0.26	0.00
Financial Services	0.05	0.00	0.12	0.02	0.01	0.20	0.07	0.12	0.01	0.41	0.00
Real Estate Services	0.03	0.00	0.11	0.01	0.02	0.27	0.04	0.04	0.06	0.41	0.00
Business Services	0.07	0.00	0.13	0.01	0.01	0.23	0.09	0.05	0.00	0.40	0.00
Personal Services	0.06	0.00	0.15	0.03	0.02	0.21	0.07	0.11	0.00	0.33	0.01

Table A9: Input–output weights by Upstream firms as buyers (uniform–pricing lens)

Buyer / Seller	Agr.	Min.	Man.	Uti.	Cons.	R. & W.	T. & ICTs	F. Serv.	RE. Serv.	B. Serv.	P. Serv.
Agriculture	0.26	0.00	0.12	0.02	0.04	0.29	0.10	0.06	0.00	0.10	0.00
Mining	0.01	0.07	0.39	0.05	0.06	0.13	0.11	0.03	0.00	0.15	0.00
Manufacturing	0.08	0.02	0.49	0.03	0.02	0.15	0.09	0.02	0.00	0.10	0.00
Utilities	0.06	0.02	0.18	0.07	0.03	0.18	0.15	0.04	0.00	0.27	0.00
Construction	0.07	0.00	0.14	0.03	0.30	0.18	0.12	0.03	0.00	0.13	0.00
Retail and Wholesale	0.12	0.01	0.27	0.01	0.02	0.38	0.07	0.03	0.00	0.10	0.00
Transport and ICTs	0.06	0.02	0.14	0.02	0.04	0.21	0.22	0.03	0.00	0.26	0.00
Financial Services	0.05	0.00	0.12	0.02	0.01	0.20	0.07	0.12	0.01	0.41	0.00
Real Estate Services	0.03	0.00	0.11	0.01	0.02	0.27	0.04	0.04	0.06	0.41	0.00
Business Services	0.07	0.00	0.13	0.01	0.01	0.23	0.09	0.05	0.00	0.40	0.00
Personal Services	0.06	0.00	0.15	0.03	0.02	0.21	0.07	0.11	0.00	0.33	0.01

We also estimate Cobb–Douglas output weights θ_s across retail sectors. Because retail–to–final transactions are linear in prices, observed retail revenues identify sectoral expenditure shares. Within each year we restrict to retailers above the 75th percentile of revenue (nonlinear–pricing lens), compute firm–level revenue shares, average within sector, and then average over 2005–2022; for the uniform–pricing lens, we repeat the construction on the full firm population. The table below report θ_s under each lens.

Table A10: Cobb–Douglas output weights by retail sector: nonlinear vs. uniform lenses

Sector	Nonlinear lens θ_s	Uniform lens θ_s
Agriculture	0.0398	0.0446
Mining	0.0119	0.0085
Manufacturing	0.1407	0.1318
Utilities	0.0681	0.0505
Construction	0.1210	0.1521
Retail and Wholesale	0.3289	0.2768
Transport and ICTs	0.0802	0.0979
Financial Services	0.0719	0.1132
Real Estate Services	0.0180	0.0152
Business Services	0.0897	0.0911
Personal Services	0.0300	0.0183

Differences across lenses are modest and reflect composition. The nonlinear–pricing lens emphasizes large–buyer moments (flat–fee dilution), while the uniform–pricing lens uses the full firm population and treats per–unit prices as marginal across buyers; for output weights, this shifts mass toward sectors dominated by large retailers under the nonlinear lens and toward sectors with many smaller retailers under the uniform lens.

D.3 Upstream Materials Elasticity of Substitution

We estimate the upstream elasticity of substitution $\sigma_{u'}$ within seller sector u', exploiting the one-off, municipality-level cost shock from Chile's March 2020 COVID-19 lockdowns. March 2020 marks the first registered COVID-19 cases in Chile, making the shock unexpected. Figure A2 maps the spatial heterogeneity of early lockdowns (municipalities in red were under lockdown in March 2020).

Figure A2: Distribution of early COVID-19 lockdowns in Chile

In our model, buyers aggregate upstream inputs from sector u' with CES. For buyer (i, s), let u^* denote the largest pre-shock supplier (by 2019 expenditure). For any $u \in \mathcal{U}_{u'}$:

$$\log\left(\frac{m_{isut}}{m_{isu^*t}}\right) = -\sigma_{u'} \log\left(\frac{p_{isut}}{p_{isu^*t}}\right),$$

where p_{isut} and m_{isut} are the buyer-facing unit price and quantity for input u at time t.

Data come from the Chilean Internal Revenue Service (SII): monthly firm-to-firm transactions, 2019–2021, with product descriptions, quantities, prices, firm identifiers, and locations. We match buyers to upstream suppliers, observe the universe of intermediate-input purchases, and track geographic lockdown exposure. For each (i, s) we identify u^* as the 2019 top supplier by value.

Define the instrument $Z_{is} = 1$ if u^* is located in a municipality under lockdown in March 2020, and 0 otherwise. This delivers a plausibly exogenous increase in the marginal cost (and price) of u^* relative to other inputs, inducing substitution away from u^*. To ensure unit prices reflect marginal prices, we restrict the estimation sample to large buyers (above the 75th percentile of average annual sales in 2019–2021). The exclusion is that the shock affects the buyer only through u^*'s relative cost. To mitigate alternative channels, we impose:

1. Buyer location: the buyer is not in a locked municipality.
2. Client base: the buyer's customers are not in locked municipalities.
3. Input scope: no other input used by the buyer was sourced from a locked municipality.

We use 12-month log differences in relative prices and quantities to remove time-invariant buyer–supplier heterogeneity and seasonality, and include buyer 6-digit sector fixed effects γ_s to absorb sector-level shocks common to all inputs. Standard errors are clustered at the buyer level.

Under the nonlinear-pricing lens, two-part tariffs imply that average unit prices converge to the marginal price as quantity rises; we therefore identify $\sigma_{u'}$ from large-buyer moments where flat-fee components are diluted, using the instrumented change in the top supplier's relative price. Under the uniform-pricing lens, per-unit prices are treated as marginal and common across buyers within seller–product–time cells; we re-estimate the same 2SLS specification on the full firm population (subject to the three exclusion conditions above). The instrument is identical across lenses; differences in estimates reflect sample composition (large-buyer restriction versus full population) and the pricing interpretation (average-versus-marginal reconciliation under nonlinear pricing versus direct marginal interpretation under uniform pricing).

For each seller sector u', we estimate a separate 2SLS on pairs $(i,s) \times u \in \mathcal{U}_{u'}$:

$$\underbrace{\Delta_{12} \log\left(\frac{p_{isut}}{p_{isu^*t}}\right)}_{\text{relative price change}} = \beta_0 + \beta_1 Z_{is} + \gamma_s + \nu_{isut}, \qquad \underbrace{\Delta_{12} \log\left(\frac{m_{isut}}{m_{isu^*t}}\right)}_{\text{relative quantity change}} = \beta_{u'} \, \Delta_{12} \widehat{\log\left(\frac{p_{isut}}{p_{isu^*t}}\right)} + \gamma_s + \varepsilon_{isut},$$

where $\beta_{u'} = -\sigma_{u'}$. The 12-month horizon targets the medium-run substitution relevant for counterfactuals.

Table A11 reports sector-specific estimates $\hat{\sigma}_{u'}$ with standard errors, first-stage F statistics, and observation counts for both lenses. Three sectors (Mining, Utilities, Real Estate Services) have insufficient post-restriction variation or yield $\hat{\sigma}_{u'} < 1$; we set their elasticities to the minimum estimated elasticity for other sectors, $\sigma = 1.44$ for nonlinear pricing lenses and $\sigma = 1.71$ for uniform price lenses. They are omitted from the table for brevity but included in all counterfactual calculations.

Table A11: Estimated elasticities of substitution by seller sector: nonlinear vs. uniform lenses

	Nonlinear lens				Uniform lens			
Sector	σ	SE	F	Obs.	σ	SE	F	Obs.
Agriculture	4.14	1.57	9.38	3,129	2.59	1.35	10.24	4,387
Manufacturing	5.19	1.14	43.10	218,027	3.89	0.67	57.68	255,462
Construction	1.44	0.43	7.36	4,725	1.71	0.42	15.21	6,111
Retail and Wholesale	3.80	0.39	94.08	680,985	5.22	0.34	236.44	953,073
Transport and ICTs	5.07	2.22	25.19	24,054	4.35	1.59	58.22	46,637
Financial Services	3.09	1.56	12.35	3,631	2.56	0.37	19.44	4,374
Business Services	6.83	1.59	14.85	3,709	6.81	1.50	27.55	5,008
Personal Services	8.08	3.24	10.68	5,255	6.21	2.37	18.29	7,728

Notes: SE = standard error; F = first-stage F statistic; Obs. = observation count in the estimation sample. Sectors omitted from the table (Mining, Utilities, Real Estate Services) are included in all counterfactuals with σ set to minimum estimated elasticity for other sectors.

Across sectors, the estimated elasticities span 1.44–8.08 under the nonlinear lens and 1.71–6.81 under the uniform lens in this sample. The nonlinear lens yields higher σ in Agriculture (4.14 vs. 2.59), Manufacturing (5.19 vs. 3.89), Transport and ICTs (5.07 vs. 4.35), Financial Services (3.09 vs. 2.56), and Personal Services (8.08 vs. 6.21), implying greater scope for substitution across upstream varieties in these sectors relative to the uniform lens. The uniform lens is higher in Construction (1.71 vs. 1.44) and Retail and Wholesale (5.22 vs. 3.80), implying easier substitution there under the uniform interpretation, while Business Services is essentially the same across lenses (6.83 vs. 6.81). In levels, the largest elasticities are in Personal Services and Business Services, indicating the most scope for substitution among the sectors shown, and the smallest are in Construction under both lenses.

D.4 Exit Rates

We estimate annual exit rates $\delta_{s\ell}$ from Chilean administrative microdata independently of any pricing assumptions. Because exit is a demographic object (presence of firm i in $t+1$), its measurement does not rely on prices or markups; accordingly, we use the same calibration under the nonlinear–pricing and uniform–pricing lenses. We consider a one–year period and a panel of firms indexed by i, years by t, sectors by s (6–digit), and firm types $\ell \in \{\text{upstream}, \text{retail}\}$ over $\mathcal{T} = \{2005, \ldots, 2022\}$. For each firm–year (i, t) define one–year survival:

$$\text{surv}_{i,t} := \mathbb{1}\{\exists \text{ an observation of firm } i \text{ in year } t+1\}.$$

At the sector–type–year level, let $N_{s_\ell,t}$ be the number of active firms and $\text{survivors}_{s_\ell,t} = \sum_{i\in(s_\ell,t)} \text{surv}_{i,t}$. The exit rate in year t is:

$$\delta_{s_\ell,t} \;=\; 1 - \frac{\text{survivors}_{s_\ell,t}}{N_{s_\ell,t}}, \qquad \delta_{s_\ell} \;=\; \frac{1}{|\mathcal{T}|}\sum_{t\in\mathcal{T}} \delta_{s_\ell,t}.$$

We aggregate 6–digit estimates to the model's 11 sectors by simple averaging within sector and report 1–digit means by firm type in Table A12. These rates are held fixed across pricing lenses in all counterfactuals.

Table A12: Exit rates (δ) by sector (means)

Sector	Retailers	Upstream	Sector mean
Agriculture	0.090	0.086	0.088
Mining	0.084	0.093	0.088
Manufacturing	0.093	0.071	0.082
Utilities	0.070	0.064	0.067
Construction	0.140	0.110	0.125
Retail and Wholesale	0.103	0.076	0.089
Transport and ICTs	0.088	0.093	0.091
Financial Services	0.101	0.062	0.081
Real Estate Services	0.115	0.099	0.107
Business Services	0.099	0.077	0.088
Personal Services	0.093	0.090	0.092
Type mean	0.098	0.084	0.091

D.5 Entry Costs

The entry cost c_{e,s_ℓ} measured in units of yearly firm-level wages. The calibration exploits the availability of firm-level accounting profits and wages. We consider a one-year period and a panel of firms indexed by i, years by t, sectors by s, and firm types by $\ell \in \{\text{upstream}, \text{retail}\}$. We compute c_{e,s_ℓ} at 6-digit sector granularity (626 sectors). Let $\mathcal{T} = \{2005, \ldots, 2022\}$ denote the estimation window.

We have access to data on firm-level yearly revenue, labor headcounts, wage-bill expenditure, material expenditure, and capital stock. We build the real user cost of capital using publicly available data.[32] Using these data, we construct yearly firm-level profits $\Pi_{i,t}$ for 2005–2022 and

[32]We use the 10-year government bond interest rate minus expected inflation plus the external financing premium.

compute the average wage per worker-year:

$$w_{i,t} \equiv \frac{\text{wagebill}_{i,t}}{\text{employees}_{i,t}}.$$

We define the per-active-firm annual profit in (s_ℓ, t) as:

$$\bar{\Pi}_{s_\ell,t} = \frac{1}{N_{s_\ell,t}} \sum_{i \in (s_\ell,t)} \Pi_{i,t}, \qquad N_{s_\ell,t} = \text{active}_{s_\ell,t}.$$

Let the sector–firm type wage per worker-year be the headcount-weighted mean:

$$w_{s_\ell,t} = \frac{\sum_{i \in (s_\ell,t)} w_{i,t} \cdot \text{employees}_{i,t}}{\sum_{i \in (s_\ell,t)} \text{employees}_{i,t}}.$$

We set w_{s_ℓ} and $\bar{\Pi}_{s_\ell}$ as averages over $\mathcal{T}$.

In our model, flat fees can redistribute profits across firm types. Calibration must therefore use the profit that accrues to the owner of the firm that pays entry. Under steady state with i.i.d. per-period profits and exogenous exit rate δ_{s_ℓ}, the expected present value (PV) of a surviving firm is:

$$\text{PV}_{s_\ell} = \frac{\bar{\Pi}_{s_\ell}}{1 - \beta(1 - \delta_{s_\ell})},$$

where $\beta = 1/(1+r)$ is the annual real discount factor and r is the annual real rate. As only a fraction $p^{\text{succ}}_{s_\ell} \in (0, 1]$ of firms have positive profits, we set $p^{\text{succ}}_{s_\ell}$ to the share of positive-profit firms in (s_ℓ). The free-entry condition is:

$$w_{s_\ell} c_{e,s_\ell} = p^{\text{succ}}_{s_\ell} \cdot \text{PV}_{s_\ell} \quad \Longrightarrow \quad c_{e,s_\ell} = \frac{p^{\text{succ}}_{s_\ell}}{w_{s_\ell}} \cdot \frac{\bar{\Pi}_{s_\ell}}{1 - \beta(1 - \delta_{s_\ell})}.$$

We implement this formula under both evaluation lenses. The exit rates δ_{s_ℓ} and wage measures w_{s_ℓ} are lens–invariant; differences across lenses arise from the lens–specific mapping of observed prices into profits and the implied positive–profit share $p^{\text{succ}}_{s_\ell}$. Under the uniform–pricing lens, per–unit prices are marginal and common across buyers within seller–product–time cells, so observed accounting profits already reflect allocative margins with no fixed–fee component; hence all firms are informative for estimating $\bar{\Pi}_{s_\ell}$ and $p^{\text{succ}}_{s_\ell}$. Under the nonlinear–pricing lens, observed average unit prices bundle marginal prices with flat fees, so for small buyers the fee–per–unit F_i/Q_i is large and contaminates variable costs and revenues, biasing profits. Restricting to large buyers makes $F_i/Q_i \approx 0$, bringing observed unit prices close to marginal and delivering lens–consistent

We use the capital depreciation rate from the LA-KLEMS database. For reference, the average government bond interest rate over 2005–2022 is 5.74%, expected inflation is 4.6%, the external financing premium is 110 basis points, and the average capital depreciation rate is 5%.

profits for the entry–cost calculation.

We compute entry costs by firm type and 6-digit sector. Table A13 reports 1–digit sector averages for retailers and upstream firms under both lenses. We also report wage bill equivalents (multiples of the annual wage bill) defined as c_{e,s_ℓ} divided by the average annual wage bill of the sector–type (average wage per worker times average employment), averaged over 2005–2022. For example, for retailers in Agriculture, the entry cost equals 3.60 annual wage bills for the average firm in that sector.

Table A13: Entry costs and wage–bill equivalents by sector: nonlinear vs. uniform lenses

	Nonlinear lens				Uniform lens			
	Retailers		Upstream		Retailers		Upstream	
Sector	c_e	Wage–bill eq.	c_e	Wage–bill eq.	c_e	Wage–bill eq.	c_e	Wage–bill eq.
Agriculture	147	3.60	137	4.65	35	3.11	40	4.07
Mining	59769	33.87	384	7.38	13410	56.28	85	6.94
Manufacturing	238	4.51	218	4.11	39	4.01	66	5.00
Utilities	2012	16.31	723	6.04	251	12.87	139	5.75
Construction	223	7.78	192	3.78	47	6.53	57	4.35
Retail and Wholesale	113	5.92	141	4.71	31	5.52	46	5.55
Transport and ICTs	667	9.89	177	5.89	140	11.34	53	6.62
Financial Services	945	11.42	616	10.02	68	5.71	107	7.61
Real Estate Services	159	14.01	213	10.23	42	9.41	46	6.92
Business Services	138	5.05	224	2.63	49	6.58	69	3.74
Personal Services	253	4.31	204	4.68	67	4.57	48	4.65

Notes: "Wage–bill eq." reports multiples of the annual wage bill for the corresponding sector–type. Sector classification follows the model's 11-sector aggregation. Exit rates and wage measures are common across lenses; profits and the positive–profit share are lens–specific as described in the text.

The mapping from profits to entry costs we use is dimensionally consistent (output units into labor units via w_{s_ℓ}) and directly comparable across sector–firm types. Compared to alternatives (e.g., inferring c_e from net-entry rates and size distributions or from structural Markov dynamics), this method is empirically simple, requires fewer auxiliary moments, and allows clean sectoral heterogeneity through the observed $\bar{\Pi}_{s_\ell}$ and w_{s_ℓ}.

D.6 Pareto Productivity Tails

The Pareto productivity tail only applies for the nonlinear price lens model. Let $\ell \in \{u, r\}$ index the firm type (upstream, retail) and $s \in \{1, \ldots, 12\}$ index 1-digit sectors within each firm type. For firm

i in (ℓ, s), let $L_i^{\ell s}$ be its number of workers. We assume a Pareto tail for $L^{\ell s}$:

$$\Pr\left(L^{\ell s} > l\right) \;=\; \left(\frac{L_{\min}^{\ell s}}{l}\right)^{\nu_s^\ell}, \qquad l \geq L_{\min}^{\ell s}, \quad \nu_s^\ell > 0,$$

equivalently, the density is $f_{L^{\ell s}}(l) = \nu_s^\ell \left(L_{\min}^{\ell s}\right)^{\nu_s^\ell} l^{-(\nu_s^\ell+1)}$. Given a threshold $L_{\min}^{\ell s}$ (baseline: firms with at least two employees), the closed-form MLE of the survival exponent is

$$\widehat{\nu}_s^\ell \;=\; \frac{n_{\ell s}}{\displaystyle\sum_{i:\, L_i^{\ell s} \geq L_{\min}^{\ell s}} \ln\left(\frac{L_i^{\ell s}}{L_{\min}^{\ell s}}\right)}, \qquad \mathrm{SE}(\widehat{\nu}_s^\ell) \approx \frac{\widehat{\nu}_s^\ell}{\sqrt{n_{\ell s}}},$$

where $n_{\ell s}$ is the number of tail observations; the number of firms in ℓ (upstream or retail) and sector s with employment at or above the threshold. We estimate ν_s^ℓ at the 1-digit sector for each layer and then map to productivity tails via the model's $l(z)$.

The model implies labor demand at productivity z:

$$l(z) \;=\; l(\tilde{z})\left(\frac{z}{\tilde{z}}\right)^{\sigma-1}, \qquad \sigma > 1,$$

which is strictly increasing in z. Fix a firm type–sector pair (ℓ, s). Suppose productivity $Z^{\ell s}$ has a Pareto upper tail with survival exponent $\kappa_s^\ell > 0$:

$$\Pr\left(Z^{\ell s} > z\right) \;=\; \left(\frac{z_{\min}^{\ell s}}{z}\right)^{\kappa_s^\ell} \quad \text{for } z \geq z_{\min}^{\ell s}.$$

Let $L^{\ell s} = l(z^{\ell s})$ with $l(z) = l(\tilde{z})\,(z/\tilde{z})^{\sigma-1}$ and define $L_{\min}^{\ell s} = l(z_{\min}^{\ell s})$. Then $L^{\ell s}$ has a Pareto upper tail with survival exponent

$$\nu_s^\ell \;=\; \frac{\kappa_s^\ell}{\sigma - 1}, \qquad \Longleftrightarrow \qquad \kappa_s^\ell \;=\; (\sigma-1)\,\nu_s^\ell.$$

Therefore, given a labor Pareto tail ν_s^ℓ and the seller-sector elasticity of substitution $\sigma_{u'}$ for sector s, the productivity Pareto tail is pinned down by $\kappa_s^u = (\sigma_{u'} - 1)\nu_s^u$ for upstream firms and $\kappa_s^r = (\varphi_{s'} - 1)\nu_s^r$ for retailers firms . We estimate labor Pareto tails using the MLE above and show them in Table A14, together with the implied productivity tails using the estimated elasticities of substitution by sector.

Table A14: Labor and Implied Productivity Pareto Tails by Sector

	Retailers		Upstream	
Sector	ν_r	$\kappa_r = (\varphi_s - 1)\nu_r$	ν_u	$\kappa_u = (\sigma_{u'} - 1)\nu_u$
Agriculture	2.49	8.82	2.63	4.18
Mining	1.43	2.40	2.20	0.99
Manufacturing	2.66	8.58	2.15	5.18
Utilities	2.17	6.38	1.94	0.87
Construction	3.23	5.13	2.19	0.99
Retail and Wholesale	3.45	24.74	2.40	6.72
Transport and ICTs	2.20	2.32	3.04	12.37
Financial Services	2.55	1.02	2.26	4.72
Real Estate Services	4.36	3.59	3.03	1.36
Business Services	2.45	4.25	1.93	8.13
Personal Services	2.03	3.17	2.58	14.69

Notes: $\kappa = (\sigma_{u'} - 1)\nu$ uses seller-sector elasticities $\sigma_{u'}$.

D.7 Final-Consumer Elasticities of Substitution

We recover the elasticity of substitution faced by the representative final consumer across retail varieties within each retail sector s_r, denoted $\varphi_{s_r} > 1$. Under the CES aggregator, the representative consumer allocates expenditure across differentiated retail varieties $j \in \mathcal{J}_{s_r}$ via

$$Q_{s_r} = \left(\sum_{j \in \mathcal{J}_{s_r}} q_j^{\frac{\varphi_{s_r}-1}{\varphi_{s_r}}} \right)^{\frac{\varphi_{s_r}}{\varphi_{s_r}-1}},$$

so the demand for variety j is isoelastic with elasticity φ_{s_r}. With linear pricing and monopolistic competition, the first-order condition yields the standard markup:

$$\mu_{s_r} \equiv \frac{p_j}{c_j} = \frac{\varphi_{s_r}}{\varphi_{s_r} - 1},$$

implying that the variable-profit share of revenue equals $1/\varphi_{s_r}$. Hence, for retailer j,

$$\Pi_j^{\text{var}} = \frac{1}{\varphi_{s_r}} R_j,$$

where R_j is sales revenue. Retailers also have to pay flat fees to upstream firms $F_{j,t}$, so that accounting profits are

$$\Pi_{j,t} \;=\; \frac{1}{\varphi_{s_r}} R_{j,t} \;-\; F_{j,t}.$$

Aggregating within sector s_r and year t gives:

$$\sum_{j \in \mathcal{J}_{s_r}} \Pi_{j,t} \;=\; \frac{1}{\varphi_{s_r,t}} \sum_{j} R_{j,t} \;-\; \sum_{j} F_{j,t},$$

which rearranges to the sector-year estimator:

$$\varphi_{s_r,t} \;=\; \frac{\sum_j R_{j,t}}{\sum_j F_{j,t} \;+\; \sum_j \Pi_{j,t}}.$$

Households face linear per–unit pricing, so the CES markup identity and the estimator hold under both lenses, nonlinear and uniform prices. Differences arise from how sectoral profits and flat fees aggregates are formed: under the nonlinear–pricing lens we compute moments on large retailers (above the 75th percentile of annual sales within each sector–year) so upstream flat–fee components do not contaminate variable cost and profits (average prices for large buyers approximate marginal prices); under the uniform–pricing lens per–unit prices are treated as marginal and common within seller–product–time cells, so we use the full retailer population.

Table A15 presents the resulting estimates of yearly averages for the 205-2022 period of φ_{s_r} by 1-digit retail sector.

Table A15: Final–consumer elasticities of substitution by retail sector: nonlinear vs. uniform lenses

Sector	Nonlinear lens φ	Uniform lens φ
Agriculture	4.82	4.54
Mining	2.66	2.68
Manufacturing	4.17	4.22
Utilities	3.95	3.94
Construction	2.61	2.59
Retail and Wholesale	8.51	8.17
Transport and ICTs	2.04	2.05
Financial Services	1.39	1.40
Real Estate Services	1.85	1.82
Business Services	2.84	2.73
Personal Services	2.62	2.56
Type mean	3.41	3.34

Notes: φ_{s_r} is computed from pooled sectoral sums of revenue, fixed costs (in labor units), and profits. The formula follows directly from the CES markup identity under linear pricing.

This approach recovers consumer-side elasticities from sectoral accounting identities under the structural model and requires no additional demand shifters or instruments.

E Quantification Material

E.1 Exposures to Final Consumption

This appendix reports the two exposure measures used to interpret the sectoral welfare opening. "SC share" is the sector's share in total supply–chain transaction value (not its share in final demand). The retail to upstream exposure indicates how strongly a marginal dollar of consumer spending reaches each upstream sector through retail purchases. The full upstream exposure lets that dollar continue circulating as upstream sectors buy from one another, capturing the total knock–on demand that ultimately lands in each upstream sector. For each measure we also report a normalized "Share" that sums to one across sectors, making the entries directly comparable.

Table A16: Final–consumption exposures by sector (levels and shares; three decimals)

Sector [SC share]	retail to upstream	Share Share(r to u)	full upstream	Share (full upstream)
Retail and Wholesale [32%]	0.170	0.319	0.462	0.364
Manufacturing [15%]	0.099	0.186	0.222	0.175
Transport and ICTs [10%]	0.051	0.096	0.172	0.136
Construction [8%]	0.065	0.121	0.107	0.084
Financial Services [18%]	0.046	0.087	0.090	0.071
Business Services [7%]	0.036	0.067	0.083	0.065
Agriculture [2%]	0.028	0.053	0.064	0.050
Utilities [3%]	0.019	0.036	0.031	0.024
Real Estate Services [1%]	0.010	0.018	0.017	0.013
Personal Services [2%]	0.006	0.012	0.011	0.009

Notes: "SC share" is the sector's share of supply–chain transaction value. The retail to upstream column tracks the immediate flow of consumer spending to upstream sectors via retail; the full upstream column adds all upstream–to–upstream rounds. Each "Share" column normalizes its exposure vector to sum to one. The mining sector (1%) has very small exposure in our data and is omitted for brevity.

The exposure map aligns closely with the sectoral welfare opening between nonlinear and linear pricing. Retail & Wholesale, Manufacturing, Transport/ICTs, and Construction carry the largest flows of final demand into the upstream network, so compressing markups where these exposures are high has the biggest aggregate payoff. Sectors with low exposures, such as Real Estate and Personal Services, contribute little to the aggregate gap even when their own wedges or masses move.